MEASUREMENT OF BOOSTED DIFFERENTIAL TOP QUARK PAIR PRODUCTION CROSS SECTION AT 13 TEV WITH THE CMS DETECTOR

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Susan J Dittmer

December 2017

MEASUREMENT OF BOOSTED DIFFERENTIAL TOP QUARK PAIR PRODUCTION CROSS SECTION AT 13 TEV WITH THE CMS DETECTOR

Susan J Dittmer, Ph.D.

Cornell University 2017

This thesis describes a measurement of the differential top quark pair production cross section for top quarks with high transverse momentum (p_T) in proton-proton collisions at a center-of-mass energy of 13 TeV. The measurement is performed using data collected by the CMS experiment during the 2016 data-taking period, corresponding to an integrated luminosity of 35.9 fb^{-1}. The measurement is performed using events where one top quark decays hadronically and is reconstructed as a single large-radius t jet with $p_T > 400$ GeV, while the other top quark decays leptonically to a b jet, an electron or a muon, and a neutrino. Jet substructure techniques are used to identify the large-radius jet as a t jet candidate. The measurement demonstrates that current simulations overpredict the rate of top quark pair production at high p_T, providing valuable feedback on the accuracy of current simulation.

BIOGRAPHICAL SKETCH

Susan Dittmer graduated from the Illinois Mathematics and Science Academy in 2007. She was accepted into the California Institute of Technology and received a B.S. in Physics with honors in 2011. She is currently completing a PhD in Physics at Cornell University.

In honor of Dr. Lucinda Dittmer

ACKNOWLEDGEMENTS

In my time at Cornell and as a member of the CMS collaboration, I have benefited greatly from an environment heavily focused on collaboration. This thesis would not be possible without support and assistance from a number of people.

First and foremost, thank you to Julia Thom-Levy for being such an encouraging, invested, and inspiring advisor. Thank you for pushing and for being patient, for being constantly involved and available, and for showing me how to mature as a researcher. I am in awe of your dedication to both research and pedagogy, and can only hope to emulate your example both in this and in maintaining a healthy work-life balance. Thank you for everything you have taught.

Thank you also to my committee members, Peter Wittich and Maxim Perelstein. Thank you for your guidance in exploring further ramifications and applications of this research, as well as for your non-committee-related teaching and advice. Thank you also for your patience and flexibility in matters of logistics.

Thank you to all the professors in the CU HEP group, including Jim Alexander and Anders Ryd. You have all made Cornell a welcoming and challenging place to conduct research in high-energy physics. Thank you for your efforts to keep Cornell a central member of the CMS collaboration and for your instruction in the classroom, as well as for your willingness to engage, career advice and research anecdotes.

Thank you for all the support provided by the Cornell Physics Department. Thank you for excellent teaching, availability of student resources, and a dedication to a tolerant and engaging place of learning. Thank you especially to Kacey Acquilano and Deb Hatfield, for your patience in dealing with absentminded graduate students and for all of the effort you put into making the department run smoothly.

Thank you to the National Science Foundation for providing funding support through my graduate career.

Thank you to Louise Skinnari, for being an amazing mentor and research partner. Thank you for teaching me to thoroughly investigate an idea and to clearly and concisely present my findings. You've been an invaluable resource, always ready to hash out an idea, answer a technical question, or help with an analysis task. Thank you for being an inspiring example of a young researcher, and always encouraging me to do better.

Thank you to Jordan Tucker, for introducing me to CERN, teaching me about the pixel detector, and helping troubleshoot tricky coding issues. I'm grateful for the benefit of your time and expertise.

Thank you to Sal Rappoccio for creating the IHeartNY group, by encouraging a group of Data Analysis School students to continue their project as a full analysis. Thank you for introducing me to the world of boosted object analysis, and always being available to answer a technical question.

Thank you to Rachel Yohay, for introducing me to the CMS pixel detector and helping me find a niche in the FPix group. I consider you a role model in leading firmly but fairly, and an example of how to pursue an academic career. I'm also glad you made me learn to drive a manual. Thank you to Stefanos Leontsinis, for encouraging me to deepen my knowledge of pixel operations, and for being another example of leadership. Thank you to the rest of the CMS pixel group, especially Danek Kotlinski, for teaching me about the pixel detector, always being available to answer questions, and being tolerant of human error.

Thank you to CERN, the LHC, and the CMS collaboration for designing, funding, and operating such an exciting experiment. I am grateful for the ef-

fort of the many people involved in these collaborations, both in directly contributing to the body of knowledge as well as making CERN such a vibrant and accepting place to do research.

Thank you to Jen Chu, for being a great friend, study buddy, roommate, and fellow researcher. I've enjoyed being able to talk with you about everything from pixel DAQ to cupcakes. Thank you to Josh Thompson, Luke Winstrom, Nathan Mirman, Dan Quach, Kevin McDermott, Margaret Zientek, Shao Min Tan, Aurelijus Rinkevicius, and Livia Soffi for making my stay at CERN so enjoyable, whether through physics talks over lunch at R1, night hikes in the Jura, or pizza dinners at Luigia's. Thank you to all of the other Cornell physics grad students, for your support and companionship in our years together.

Thank you Mom, for encouraging me to do the things I love and never letting me settle for less than my best. Your love, wisdom, and example have made me who I am today. Thank you Dad, for being my steady support – for always wanting to know about my research (even when it gets technical), encouraging me when I get frustrated, and giving me a push when I need it. I am so grateful for your support and love. Thank you to my sister Sarah, for being my number one cheerleader and such an amazing example of passion and determination. I can't wait until we're Doctors Dittmer together.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

This thesis presents a measurement of the production rate of top quark pairs in proton-proton (pp) collisions at the Large Hadron Collider (LHC). Specifically, the measurement focuses on top quarks which are produced at high energies. When these top quarks decay, the decay products are produced collimated due to the significant Lorentz boost. The decay of a top quark produces a W boson and b quark, with the W boson further decaying either to a pair of quarks or a lepton and neutrino. The $t \to Wb \to q\bar{q}b$ process is referred to as hadronic top quark decay, while the $t \to Wb \to \ell\nu b$ process is referred to as leptonic top quark decay. The final state quarks hadronize, producing showers of hadrons known as jets. At low top quark energies, the three jets of the hadronic top decay are well separated and may be individually identified; at high energies, however, they merge into a single wide top quark jet (t jet).

This thesis measures the top quark pair production cross section in the semileptonic final state, where one top quark decays leptonically and the other top quark decays hadronically. The final state therefore contains a t jet from the hadronic top quark decay, as well as a lepton and b quark jet (b jet) from the leptonic top quark decay. All final state particles are identified through interactions with the Compact Muon Solenoid (CMS) detector. This unique final state is beyond the reach of low-energy or resolved $t\bar{t}$ cross section measurements and requires a dedicated analysis.

Measurements in boosted topologies have become more important as the LHC moves to higher collision energies and collision rates. A continuous cycle of tuning simulation parameters using measurements in collision data is nec-

essary to provide accurate and precise predictions of event rates and kinematics. As the LHC reach extends to higher energies, measurements in the boosted regime are necessary to ensure these predictions remain accurate in the high energy tail. Top quark measurements are of particular importance, as many models predict the top quark to be sensitive to previously unobserved physics. Precise and accurate predictions at high energies are therefore necessary to search for any deviation from expected predictions.

This thesis is structured as follows. Chapter 2 contains a summary of the current theory of high-energy physics (called the standard model or SM), its limitations, and a brief discussion of possible extensions to the theory. Emphasis is placed on the SM predictions for the $t\bar{t}$ cross section, as well as sensitivity of the $t\bar{t}$ cross section to non-SM signals. Chapter 3 gives an overview of the LHC and the CMS detector. Chapter 4 provides a more in-depth look at the CMS pixel tracker, which is vitally important to the precise measurements of jet substructure needed to identify boosted objects. Since I was involved with operation and calibration of the pixel tracker, particular emphasis is placed on these aspects. Chapter 5 describes the general object reconstruction used by CMS to identify stable final-state particles produced in the collision, as well as the techniques used to reconstruct higher-level objects from these particles. This includes the algorithms used to identify t jets, b jets, and leptons. Finally, Chapter 6 describes the measurement of the differential $t\bar{t}$ cross section.

CHAPTER 2

THEORY

2.1 Standard Model

The field of high energy physics seeks to understand the matter and interactions which make up the universe at their most fundamental level. The current theory of high energy physics is the standard model (SM). In this section we describe the SM in the formalism of quantum field theory, following the discussions in [1,2].

The SM is a gauge theory with local $SU(3)_C \times SU(2)_L \times U(1)_Y$ symmetry, where the $SU(3)_C$ component describes quantum chromodynamics (QCD) and the $SU(2)_L \times U(1)_Y$ component describes electroweak (EWK) interactions. The EWK symmetry is spontaneously broken, $SU(2)_L \times U(1)_Y \rightarrow U(1)_{EM}$. The theory contains twelve bosons: eight G_a^μ gluons corresponding to the $SU(3)_C$ generators, three W_b^μ corresponding to the generators of $SU(2)_L$, and a B^μ corresponding to the generator of $U(1)_Y$. The $SU(3)_C$, $SU(2)_L$, and $U(1)_Y$ symmetries contribute coupling constants g_S, g, and g', respectively. The covariant derivative may be written

$$D^\mu = \delta^\mu + ig_S G_a^\mu L_a + igW_a^\mu T_a + ig'YB^\mu, \tag{2.1}$$

where L_a and T_a are the $SU(3)_C$ and $SU(2)_L$ generators, respectively. The fermionic field content consists of the left handed quarks Q_{Li}, the right handed quarks U_{Ri} and D_{Ri}, the left handed leptons L_{Li}, and the right handed leptons E_{Ri}, where $i = 1, 2, 3$ is the generation. The quarks are $SU(3)_C$ triplets, while the leptons are $SU(3)_C$ singlets. The left handed quarks and leptons are $SU(2)_L$ doublets, while the right-handed quarks and leptons are $SU(2)_L$ singlets. The

SM also contains a scalar Higgs field ϕ, which mediates the EWK spontaneous symmetry breaking (SSB). ϕ is a $SU(2)_L$ doublet. The SM Lagrangian can then be written

$$\mathcal{L} = \mathcal{L}_{kin} + \mathcal{L}_{\phi} + \mathcal{L}_{Yuk} \tag{2.2}$$

where

$$\begin{aligned} \mathcal{L}_{kin} = \quad & -\tfrac{1}{4} G_a^{\mu\nu} G_{a\mu\nu} - \tfrac{1}{4} W_b^{\mu\nu} W_{b\mu\nu} - \tfrac{1}{4} B^{\mu\nu} B_{\mu\nu} + (D^{\mu}\phi)^{\dagger} \left(D_{\mu}\phi\right) \\ & + i\bar{Q}_{Li} \not{D} Q_{Li} + i\bar{U}_{Ri} \not{D} U_{Ri} + i\bar{D}_{Ri} \not{D} D_{Ri} + i\bar{L}_{Li} \not{D} L_{Li} + i\bar{E}_{Ri} \not{D} E_{Ri} \end{aligned} \tag{2.3}$$

$$\mathcal{L}_{\phi} = -\mu^2 \phi^{\dagger}\phi - \lambda \left(\phi^{\dagger}\phi\right)^2 \tag{2.4}$$

$$\mathcal{L}_{Yuk} = Y^u_{ij} \bar{Q}_{Li} U_{Rj} \widetilde{\phi} + Y^d_{ij} \bar{Q}_{Li} D_{Rj} \phi + Y^e_{ij} \bar{L}_{Li} E_{Rj} \phi + \text{h.c.} \tag{2.5}$$

For $\mu^2 < 0$ and $\lambda > 0$, the ground state of the Higgs potential $\mathcal{L}_{\phi}$ spontaneously breaks the $SU(2)_L \times U(1)_Y$ symmetry with a vacuum expectation value of $|\langle\phi\rangle| = \sqrt{-\mu^2/2\lambda} = v/\sqrt{2}$. The SSB gives masses to the three bosons that correspond to the generators of the broken symmetry. The mass eigenstates $W^{\pm}$ and Z^0, as well as the massless A^0, are related to W_b^{μ} and B^{μ} via

$$W^{\pm} = \frac{1}{\sqrt{2}} \left(W_1 \mp i W_2\right) \tag{2.6}$$

$$Z^0 = \cos\theta_W W_3 - \sin\theta_W B \tag{2.7}$$

$$A^0 = \sin\theta_W W_3 + \cos\theta_W B \tag{2.8}$$

where the weak mixing angle θ_W relates the W and Z boson masses such that $\cos\theta_W = m_W/m_Z = g/\sqrt{g^2 + g'^2}$.

Under SSB, the fermions acquire Dirac masses $m_f = y_f v/\sqrt{2}$, where y_f comes from the diagonal Yukawa matrices $Y^u = \text{diag}(y_u, y_c, y_t)$, $Y^d = \text{diag}(y_d, y_s, y_b)$, and

$Y^e = \text{diag}(y_e, y_\mu, y_\tau)$. As y_f also parameterizes the fermion coupling to the Higgs, particles with larger mass have a stronger Higgs coupling. A summary of the SM particle content is given in Table 2.1, which lists the particle names, spin, color (QCD charge), EM charge, and mass. As this thesis concerns a measurement of the top quark cross section, a summary of quark interactions in the SM is given in Table 2.2.

Particle	Spin	Color	Charge [e]	Mass [v]
$W^\pm$	1	(1)	± 1	$\frac{1}{2}g$
Z^0	1	(1)	0	$\frac{1}{2}\sqrt{g^2+g'^2}$
A^0	1	(1)	0	0
g	1	(8)	0	0
h	0	(1)	0	$\sqrt{2\lambda}$
e, μ, τ	1/2	(1)	-1	$y_{e,\mu,\tau}/\sqrt{2}$
ν_e, ν_μ, ν_τ	1/2	(1)	0	0
u, c, t	1/2	(3)	+2/3	$y_{u,c,t}/\sqrt{2}$
d, s, b	1/2	(3)	-1/3	$y_{d,s,b}/\sqrt{2}$

Table 2.1: Particle content of the SM

Interaction	Force Carrier	Coupling
Electromagnetic	Photon (γ)	eQ
Strong	Gluon (g)	g_S
Weak (neutral current)	Z^0 boson	$\frac{g}{\cos\theta_W}\left[\left(T_3 - Q\sin^2\theta_W\right)P_L - \left(Q\sin^2\theta_W\right)P_R\right]$
Weak (charged current)	$W^\pm$ boson	$\frac{1}{\sqrt{2}}gV_{CKM}P_L$
Yukawa	Higgs boson (h)	y_q

Table 2.2: Quark interactions in the SM. Q gives the quark EM charge, V_{CKM} is the Cabibbo-Kobayashi-Maskawa matrix describing quark flavor mixing, and $P_{L,R}$ are the projection operators for the left- and right-handed quark fields.

The QCD component of the SM is of particular importance when describing pp collisions. The strong coupling $\alpha_S \equiv g_S^2/4\pi$ depends logarithmically on the energy scale, becoming smaller as the energy scale increases or length scale shrinks. This is due to the phenomenon of asymptotic freedom, where the mag-

nitude of a color charge effectively increases with increasing distance due to the color charge carried by virtual gluons produced from the vacuum. Asymptotic freedom, also known as quark confinement, prevents bare quarks from being observed in nature. Instead, quarks form color-singlet bound states known as hadrons. Quarks, as color triplets, may carry one of three color charges – red, blue, or green. Anti-quarks have anti-color charge. Color-singlet bound states may thus be formed from three quarks of red, blue, and green color (baryons) or two quarks with a given color and anti-color (mesons). The quarks and gluons contained within a hadron are known as partons. Quarks produced in particle collisions rapidly undergo a process called hadronization, where they combine with quarks and gluons pulled from the vacuum to form sprays of hadrons known as jets.

The SM predictions for physical observables are calculated through a system of successive approximations known as perturbation theory. Observables can be expressed as a power series in a given parameter, typically one of the fundamental couplings. The calculation is generally truncated after a given power term, referred to as the order of the calculation. Leading order (LO) calculations stop at the first power, next to leading order (NLO) at the second, and so forth. Terms in the power series are often represented as Feynman diagrams, which depict the interaction graphically while encoding the relevant mathematics.

2.2 New Physics

2.2.1 Motivation

The SM has been tested to a high degree of precision since its inception. To date, no evidence has been found for physics beyond the SM. However, the SM is known to be an incomplete theory. One shortcoming of the SM is the precise cancellations needed to explain the observed Higgs mass, referred to as the hierarchy problem. The Higgs mass receives corrections at 1-loop order such as those shown in Figure 2.1 [3].

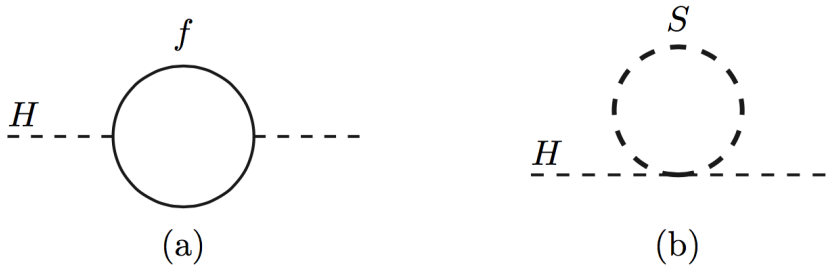


Figure 2.1: Feynman diagrams showing fermion (a) and scalar (b) 1-loop corrections to the Higgs mass [3].

The fermion loop gives a correction

$$(\Delta m_H)^2 \approx \frac{-|y_f|^2}{8\pi^2}\Lambda^2 \tag{2.9}$$

and the scalar loop a correction

$$(\Delta m_H)^2 \approx \frac{y_s}{16\pi^2}[\Lambda^2 - 2m_s ln(\frac{\Lambda}{m_s}) + ...] \tag{2.10}$$

where $y_{f,s}$ are the fermion and scalar Yukawa couplings, m_s is the scalar mass and Λ is the energy scale of new strong interactions. In order for the Higgs mass to remain light, either Λ must be very small or the various corrections to the Higgs mass must precisely cancel. The new physics scale has been probed experimentally through precision EWK measurements. These measurements impose a limit of $\Lambda > 10$ TeV [4], eliminating small Λ as a solution to the hierarchy problem. However, the fine tuning necessary to precisely cancel the various contributions to the Higgs mass is seen as unnatural by most physicists.

Another shortcoming of the SM lies in the description of gravitational interactions. In a quantum field theory framework, the gravitational force is mediated by a massless, neutral, spin 2 boson known as a graviton. The weakness of gravity prevents the graviton from being observed directly; however, this does not invalidate quantum gravity as a viable theory. More concerning, in the absence of new physics quantum gravity is a non-renormalizable theory. Renormalization describes the process used to eliminate infinities, for example those from loop diagrams, when calculating the quantum mechanical amplitude for a process. Such infinities are unphysical, essentially corresponding to a probability greater than one to observe a given process. To remove such infinities, renormalization specifies a scale dependence of model parameters such that the infinities cancel. However, it is impossible to perform this renormalization for quantum gravity, leading to infinities at high energies. The breakdown scale for quantum gravity is the Planck scale, 1.2×10^{19} GeV. In order for quantum gravity to be viable, new physics must exist below this scale. Less fundamentally, the relative weakness of gravity is also seen as a shortcoming of the SM, as the extreme weakness of gravity when compared to the other fundamental forces appears unnatural.

Finally, the SM does not explain certain observed phenomena in the universe, such as dark matter or dark energy. Dark matter refers to matter which interacts gravitationally with other matter, but interacts very weakly via the other fundamental forces. Dark matter has been observed through galactic rotation curves, gravitational lensing, and other measurements in which visible objects such as stars or dust experience larger gravitational forces than can be explained by other visible objects. Dark energy is a novel substance theorized to have repulsive self-interaction. Dark energy was proposed to explain the observed acceleration of universal expansion, despite attractive gravitational interactions between matter. Under the current cosmological model, the energy density of the universe would have to be 74% dark energy to match the observed acceleration.

On account of these shortcomings, most physicists believe the SM to be the low-energy approximation of a more complete theory. However, with no evidence of physics beyond the SM, the form of this more complete theory remains unknown.

2.2.2 Models

A wide variety of models have been proposed to extend the SM. The sections below describe supersymmetry and extra dimensions, two examples of new physics models which are well-described and relevant for top quark measurements. This section is not meant to be an exhaustive look at new physics, but rather to give an idea of several models which could influence the differential $t\bar{t}$ cross section.

Supersymmetry

Supersymmetry (SUSY) is a model introduced to solve the hierarchy problem [3]. Since fermion and boson corrections to the Higgs mass have opposite sign, introducing a fermion term for every SM boson term (and vice versa) would lead to a cancellation of Higgs mass corrections at all orders, providing a natural solution to the hierarchy problem. This symmetry between fermions and bosons is known as SUSY. The minimal supersymmetric extension to the SM (MSSM) pairs each SM particle with a corresponding superpartner, forming supermultiplets containing equal numbers of bosonic and fermionic degrees of freedom. Chiral supermultiplets contain a SM quark or lepton and a complex scalar squark or slepton superpartner. Gauge supermultiplets contain a SM vector boson and a Weyl fermion gaugino. Two Higgs supermultiplets are necessary to prevent gauge anomalies. The MSSM particle spectrum is summarized in Table 2.3.

Names		spin 0	spin 1/2	$SU(3)_C, SU(2)_L, U(1)_Y$
squarks, quarks	Q	$\left(\tilde{u}_L \;\; \tilde{d}_L\right)$	$(u_L \;\; d_L)$	$\left(3, 2, \frac{1}{6}\right)$
	$\bar{u}$	$\tilde{u}_R^*$	$u_R^\dagger$	$\left(\bar{3}, 1, -\frac{2}{3}\right)$
	$\bar{d}$	$\tilde{d}_R^*$	$d_R^\dagger$	$\left(\bar{3}, 1, \frac{1}{3}\right)$
sleptons, leptons	L	$(\tilde{\nu} \;\; \tilde{e}_L)$	$(\nu \;\; e_L)$	$\left(1, 2, -\frac{1}{2}\right)$
	$\bar{e}$	$\tilde{e}_R^*$	$e_R^\dagger$	$(1, 1, 1)$
Higgs, higgsinos	H_u	$\left(H_u^+ \;\; H_u^0\right)$	$\left(\tilde{H}_u^+ \;\; \tilde{H}_u^0\right)$	$\left(1, 2, +\frac{1}{2}\right)$
	H_d	$\left(H_d^0 \;\; H_d^-\right)$	$\left(\tilde{H}_d^0 \;\; \tilde{H}_d^-\right)$	$\left(1, 2, -\frac{1}{2}\right)$

Names	spin 1/2	spin 1	$SU(3)_C, SU(2)_L, U(1)_Y$
gluino, gluon	$\tilde{g}$	g	$(8, 1, 0)$
winos, W bosons	$\tilde{W}^\pm \;\; \tilde{W}^0$	$W^\pm \;\; W^0$	$(1, 3, 0)$
bino, B boson	$\tilde{B}^0$	B^0	$(1, 1, 0)$

Table 2.3: Particle content of the MSSM, including symmetry representations [3].

Since no new physics has been observed at SM mass scales, SUSY must be a broken symmetry, with SUSY particle masses larger than their SM partners. In order to preserve the cancellations which solve the hierarchy problem, however, SUSY breaking must be soft and the mass splittings between SM and SUSY particles not too large. The mechanism of SUSY breaking is highly debated. In the simplest case, ignorance of the true mechanism of SUSY breaking is parameterized by introducing a generic soft SUSY-breaking term to the Lagrangian. The generic nature of SUSY results in a phase space which is too large to be treated effectively. For the purpose of searches, simplified models which reduce the number of SUSY parameters are often used. Other constraints on SUSY are imposed to align the model with observations in data. R-parity conservation [5] is one such constraint, introduced to prevent baryon- and lepton-number violating processes such as proton decay. SM particles have an R-parity of 1, while their superpartners have an R-parity of -1. One consequence of R-parity conservation is that the lightest supersymmetric particle (LSP) is expected to be stable, as it has no decay mode which preserves R-parity. If the LSP is electrically neutral and weakly interacting, as is favored, it becomes a natural dark matter candidate particle.

Since the top quark has the strongest Yukawa coupling y_f to the Higgs, it contributes the largest correction to the Higgs mass. The mass splitting between the top quark and its SUSY partner, the top squark, must therefore be particularly small, of order 1 TeV or less if the theory is natural. The presumed low mass of the top squark, combined with the top squark coupling to the top quark, make searches for SUSY in $t\bar{t}$ final states particularly compelling.

Extra Dimensions

New physics models with extra dimensions explain the relative weakness of gravity by proposing a spacetime with five or more dimensions, such that gravity is weak due to its propagation in the extra dimension. A particularly compelling extra dimensional model is the bulk Randall-Sundrum (RS) model [6,7]. The RS model proposes a single warped extra dimension, definining spacetime to be a 5D anti-deSitter space (AdS_5) with metric

$$ds^2 = e^{-2kr_c\phi}\eta_{\mu\nu}dx^\mu dx^\nu - r_c^2 d\phi^2. \tag{2.11}$$

Here, x is the standard 4D coordinate, ϕ is the coordinate in the 5th dimension, r_c is the radius of curvature of the 5th dimension, and k is a scale on the order of the Planck scale. The space is bounded by 4D branes at $\phi = 0$ and $\phi = \pi$. The full 5D theory can be written as an effective theory in 4D through a Kaluza-Klein (KK) reduction, where the mass scales of the effective 4D theory are related to the 5D mass scales by $e^{-2kr_c\phi}$. Planck-scale gravity and TeV-scale weak interactions can therefore be realized with an appropriate r_c if the 5D mass scale is on the order of the Planck scale, the graviton wave function peaks near the $\phi = 0$ brane, and the EWK fields are localized at the $\phi = \pi$ brane. For this reason, the $\phi = 0$ brane is typically referred to as the Planck brane, while the $\phi = \pi$ brane is referred to as the TeV brane.

In the bulk RS model, the fermion and gauge boson fields are allowed to propagate in the AdS_5 bulk, rather than being confined to a given brane. The SM fermion mass hierarchy then arises from the localization in the bulk, rather than a more fundamental source. The large top quark Yukawa coupling can be

explained by localizing the top quark (specifically t_R) near the TeV brane. To remain consistent with precision EWK measurements, $(t,b)_L$ is flat in the bulk and the remaining fermion fields are localized near the Planck brane. A diagram of the AdS_5 space in the bulk RS scenario, showing quark localization, is shown in Figure 2.2.

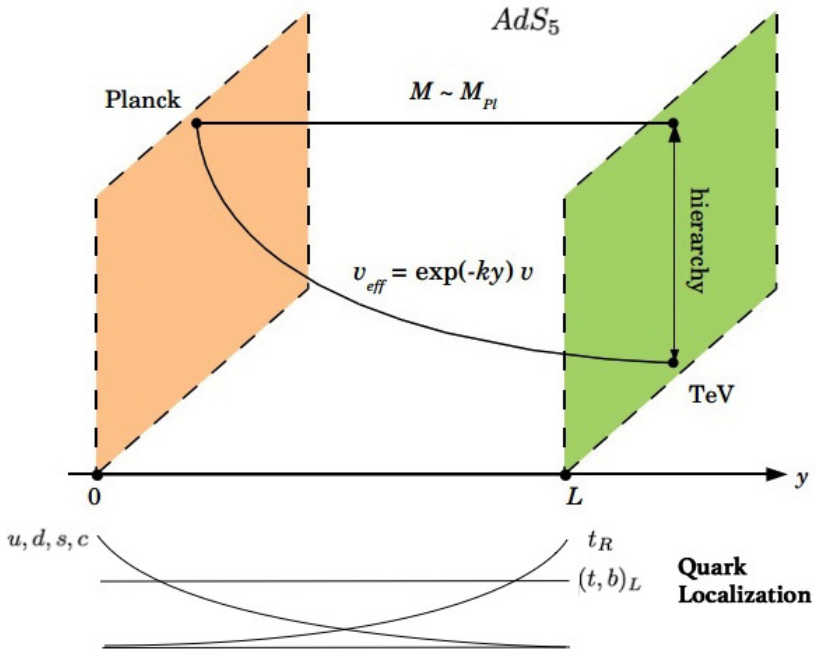


Figure 2.2: Depiction of the AdS_5 space in the bulk RS scenario. The Planck and Higgs 4D branes are located at $y = r_c\phi = 0$ and $y = r_c\pi = L$, respectively. The 4D effective scale depends exponentially on the 5D coordinate, introducing a mass hierarchy at the Higgs brane. The graviton and the Higgs boson are localized towards the Planck and Higgs branes, respectively. The top quark is localized towards the Higgs brane to explain the high top quark mass, while the lighter quarks are localized towards the Planck brane. This diagram is adapted from one found in [8].

As a consequence of the KK reduction, each particle in the 4D effective theory receives a spectrum of excited modes. The KK excited bosons are localized near the TeV brane, and consequently couple strongly to top quarks. As the most strongly coupled boson, the KK gluon (KKG) has the highest production

rate. The process KKG $\rightarrow t\bar{t}$ is therefore one of the clearest expected signals of the bulk RS model. As the KK gluon mass scale is of TeV order, the $t\bar{t}$ pair is produced boosted. The boosted $t\bar{t}$ final state is therefore particularly sensitive to bulk RS signals.

2.3 Simulation

Monte Carlo (MC) simulation is used to translate theoretical particle physics models into predictions of actual final states, as might be produced in a particle collision. MC simulation simply involves generating random numbers to sample any relevant probability distributions, then constructing an event with the sampled properties. This section describes the procedure used to simulate the results of hadron collisions, following the discussions given in [9,10].

Event generation begins with simulation of the hard interaction. The parton distribution functions (PDFs) for the colliding protons are sampled to determine the proton constituents (quarks or gluons) participating in the hard interaction and their kinematics. The hard process is then constructed by sampling appropriate cross sections, branching fractions, and kinematic distributions to determine the particles produced in the collision, the further decay of unstable particles, and the kinematics of final state particles, respectively.

After the hard process is determined, various algorithms are used to add initial and final state radiation (ISR and FSR) to the event, describing the radiation of gluons by inital and final state particles. The phase space is dominated by soft and collinear radiation of gluons. In the soft collinear approximation, parton shower (PS) models use Sudakov form factors to determine where to

split off vanishingly soft or collinear gluons from hard gluon or quark propagators. PS models account for the quark and gluon color information. To model wide-angle or hard radiation, tree-level matrix elements (MEs) are calculated to directly determine the cross section for the hard process plus a given number of additional radiated gluons. The scale of separation between the initial parton and radiated gluon must be sufficiently large for fixed-order perturbation theory to apply. The ME amplitudes are reweighted to ensure that the ME and PS methods provide a coherent description of radiation at all scales. This is referred to as ME+PS matching.

Finally, hadronization algorithms are used to convert bare final state quarks into hadron showers. Two general models are used to describe hadronization. In the string model, color strings are generated between color-connected $q\bar{q}$ pairs. These strings carry a potential depending on the quark separation. A string may be broken through the generation of a new $q\bar{q}$ pair from vacuum, forming the endpoints of the two new strings. This breaking occurs whenever the potential stored in the string is larger than the $q\bar{q}$ pair mass. The string model creates strings between all partons in the final state. These strings are successively broken, moving inwards from the edges of the event. Once all strings have been broken, nearby quarks are grouped into color-singlet hadrons. The break points are chosen such that the resultant hadrons are on-shell.

The second hadronization algorithm is the cluster model. In the cluster model, all gluons are first split into $q\bar{q}$ pairs. The ensemble of quarks is then divided into color-singlet groups known as clusters. Clusters with high invariant mass are split recursively until the cluster mass is consistent with hadron masses. The low-mass clusters are then converted directly into hadrons, while

high-mass clusters are converted into excited mesons which decay to a pair of hadrons. In this conversion process, momentum is redistributed between neighbors as necessary to produce on-shell hadrons.

In addition to the hard event, other processes contributing to the final state are modeled. The underlying event describes color interactions between bystander quarks in the primary pp collision, while pileup describes additional soft interactions between other protons in the bunch. Hadronization algorithms are also applied to the products of these interactions, where necessary.

While aspects of event simulation such as modeling the hard process rely on SM predictions, aspects such as hadronization instead utilize heuristic models developed to match phenomenology. As such, many simulation parameters have no inherent theoretical value. Instead, these parameters are chosen to provide the best description of data. Examples of such parameters include the string tension in the string hadronization model, or the ME+PS matching factors. Other simulation parameters may differ from their expected SM values in order to compensate for differences between exact SM descriptions and phenomenologically driven approximations. Examples of such parameters include the strong scales α_S used to calculate ISR and FSR. In general, tuning simulation parameters to match observations in data is essential to ensuring accurate simulation. Parameters are tuned in order of their impact on the generated final state, with parameters such as α_s tuned first. Over-tuning is avoided so as not to bias the simulation.

To account for detector acceptance and resolution, the simulated events are passed through a simulation of the detector.

2.4 Top Quark Pair Production

The top quark is of particular interest both for SM measurements and for new physics searches. Since the top quark has the strongest Yukawa coupling to the Higgs, measurements of the top quark mass in conjunction with the mass of the Higgs and other heavy bosons yield precision tests of EWK dynamics. Additionally, measurements of top properties such as spin correlation and charge asymmetry probe predictions of perturbative QCD.

This thesis presents a measurement of the $t\bar{t}$ cross section as a function of the top quark transverse momentum (p_T), for top quarks with $p_T > 400$ GeV. At the LHC, $t\bar{t}$ production proceeds primarily through gluon fusion. At leading order in perturbation theory, the Feynman diagrams contributing to the process $gg \to t\bar{t}$ are shown in Figure 2.3.

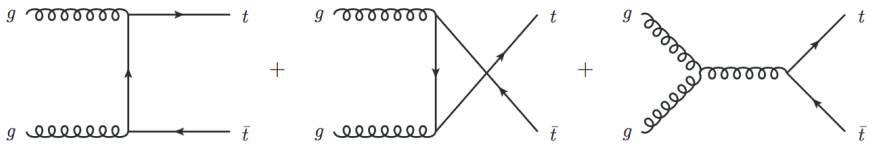


Figure 2.3: Leading order Feynman diagrams contributing to $t\bar{t}$ production at the LHC [11].

The inclusive $t\bar{t}$ cross section is sensitive to the strong coupling α_S, while the $t\bar{t}$ cross section as a function of the top quark p_T is further sensitive to the top quark mass and proton PDF. The latest theoretical calculation of the differential $t\bar{t}$ cross section as a function of top quark p_T for pp collisions at 13 TeV is shown in Figure 2.4 [12]. The prediction shown is calculated to next to next to leading order (NNLO) in QCD perturbation theory, and includes EWK corrections. Specifically, the computation includes NNLO QCD terms of $O(\alpha_S^3)$, NLO EWK effects of $O(\alpha_S^2\alpha)$, subleading NLO effects of $O(\alpha_S\alpha^2)$ and $O(\alpha^3)$, and LO

effects of $O(\alpha_S \alpha)$ and $O(\alpha^2)$. The dominant uncertainty on the theoretical prediction comes from the uncertainty on the LUXQED PDF set and the α_S scale uncertainty.

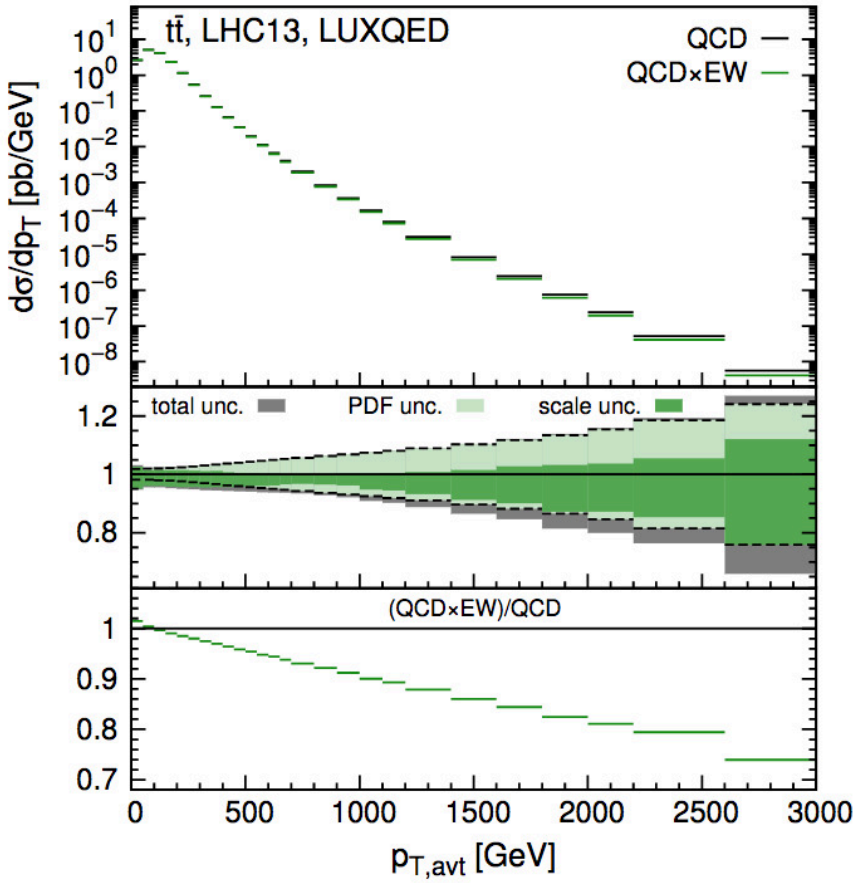


Figure 2.4: Differential $t\bar{t}$ cross section as a function of the average top quark p_T calculated to NNLO in QCD with EWK corrections [12]. The dominant uncertainty on the cross section comes from uncertainties in α_S, while the PDF uncertainty also contributes. Neglecting EWK corrections shifts the expected top quark p_T spectrum to higher values.

Measurements of the differential $t\bar{t}$ cross section are used to provide feedback on simulation tunes. In the measurement presented in this thesis, $t\bar{t}$ events are simulated using POWHEG (v.2) [13–16] for the NLO ME calculation, matched to PYTHIA8 [17–19] for the PS and hadronization. The best description of $t\bar{t}$ data

is given by the POWHEG+PYTHIA8 tune CUETP8M2T4 [20], which was developed in part by fitting measurements of $t\bar{t}$ quantities as described in [21]. Figure 2.5 demonstrates the effects of varying the ME calculation, PS algorithm, or simulation tune on the predicted $t\bar{t}$ cross section as a function of top quark p_T. The predicted $t\bar{t}$ cross sections are compared to the $t\bar{t}$ cross section as measured in 2015 data, demonstrating which methods give the best description of the data.

In addition to SM properties, top quark measurements are sensitive to new physics. Many new physics models, including extra dimensions, predict new heavy resonances which would decay to a pair of top quarks. The clearest signal of these models would be a resonant peak in the differential $t\bar{t}$ cross section with respect to the $t\bar{t}$ invariant mass. A recent search for a new heavy Z' particle with properties similar to a Z boson decaying to a $t\bar{t}$ pair was performed using 2.6 fb^{-1} of 13 TeV collision data [26]. No evidence of any resonance was observed. The Z' boson can be interpreted as a KK gluon in the bulk RS scenario. In this case, limits can be set on the production cross section of the RS KK gluon, as shown in Figure 2.6.

Alternatively, models such as the MSSM predict pair production of heavy top partners, which would then decay to a $t\bar{t}$ + X final state. Instead of a resonance, these models predict a general enhancement of the $t\bar{t}$ cross section. A recent search for top squark pair production ($\tilde{t}\tilde{t}^*$), where each top squark decays to a top quark and neutralino $\tilde{\chi}_1^0$, was performed at 13 TeV in 12.9 $\text{fb}^{-1}$of data [27]. No evidence of any excess was observed. In the context of a simplified SUSY model where the top squarks are strongly produced, the cross section for this process can be parameterized in terms of the top squark and neutralino masses. In this case, limits can be set on the range of top squark and neutralino

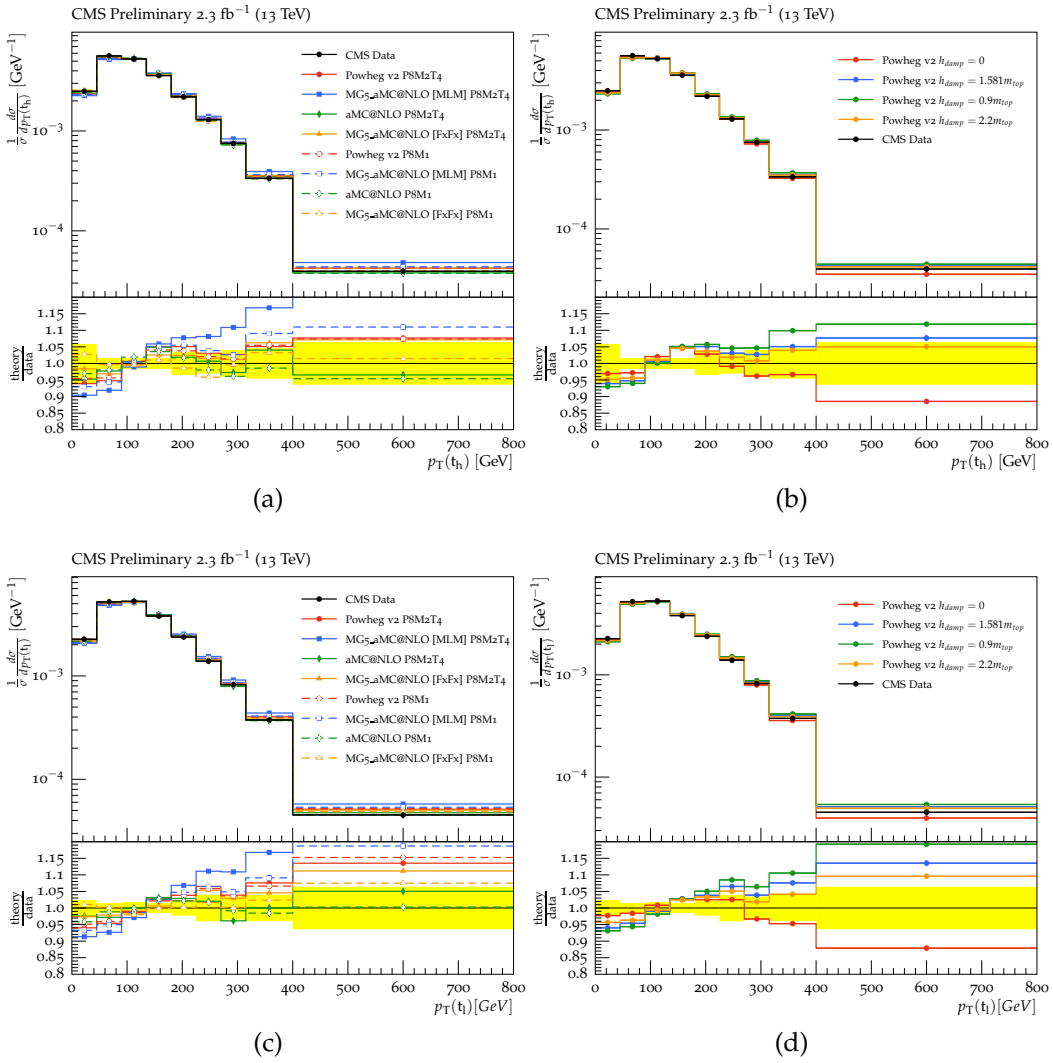


Figure 2.5: Comparison between the measured $t\bar{t}$ cross section as a function of top quark p_T and various simulated predictions. The differential $t\bar{t}$ cross section was measured at 13 TeV in the semileptonic final state, using 2.3 fb^{-1} of data collected in 2015. The top plots show the cross section as a function of the p_T of the hadronically decaying top, while the bottom plots show the cross section as a function of the p_T of the leptonically decaying top. The left plots show comparisons between the predictions given by the POWHEG v2, MG5_aMC@NLO [22], and aMC@NLO [22] ME calculations when combined with the PYTHIA8 PS using the CUETP8M1 [23] or CUETP8M2T4 tunes. The MLM [24] and FxFx [25] matching schemes are alternately used to match MG5_aMC@NLO with PYTHIA8. The right plots show comparisons between different POWHEG+PYTHIA8 predictions using tune CUETP8M2T4, where the only parameter varied is the POWHEG v2 parameter h_{damp} controlling ME+PS matching.

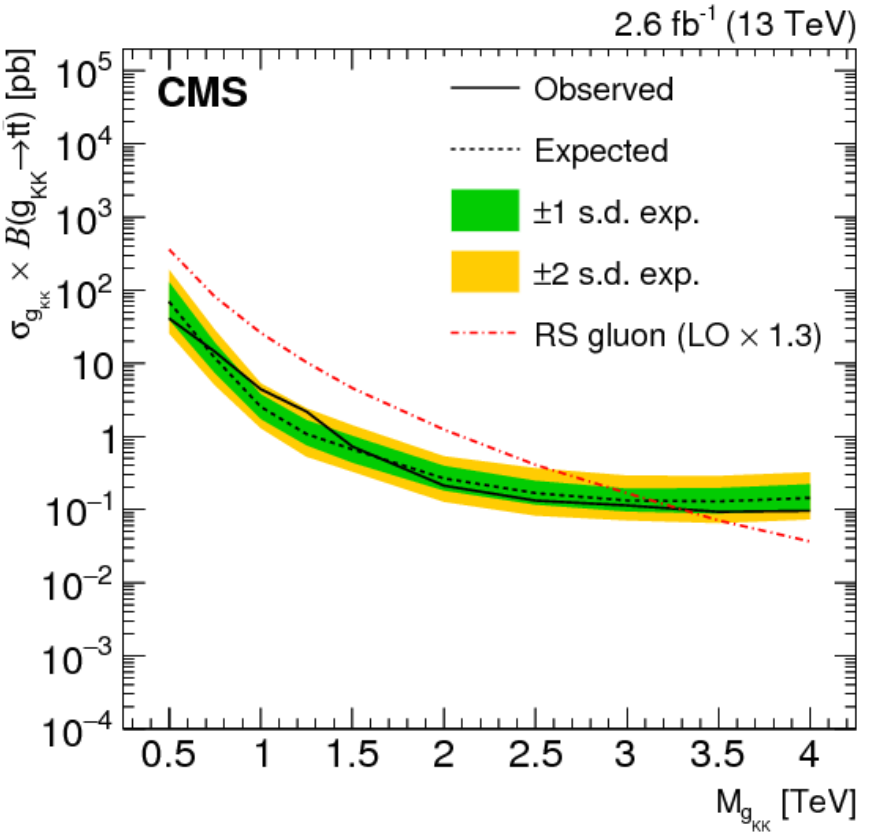


Figure 2.6: Observed and expected limits on the KK gluon cross section in the bulk RS scenario, as derived from the search for a heavy resonance in the $t\bar{t}$ final state [26]. Phase space above the observed exclusion limit (solid black line) is excluded. The expected limit describes the region of sensitivity, as determined by comparing the expected signals in the SM-only and bulk RS scenarios. The red line gives the expected cross section of the KK gluon as a function of its mass, in the bulk RS scenario.

masses, as shown in Figure 2.7.

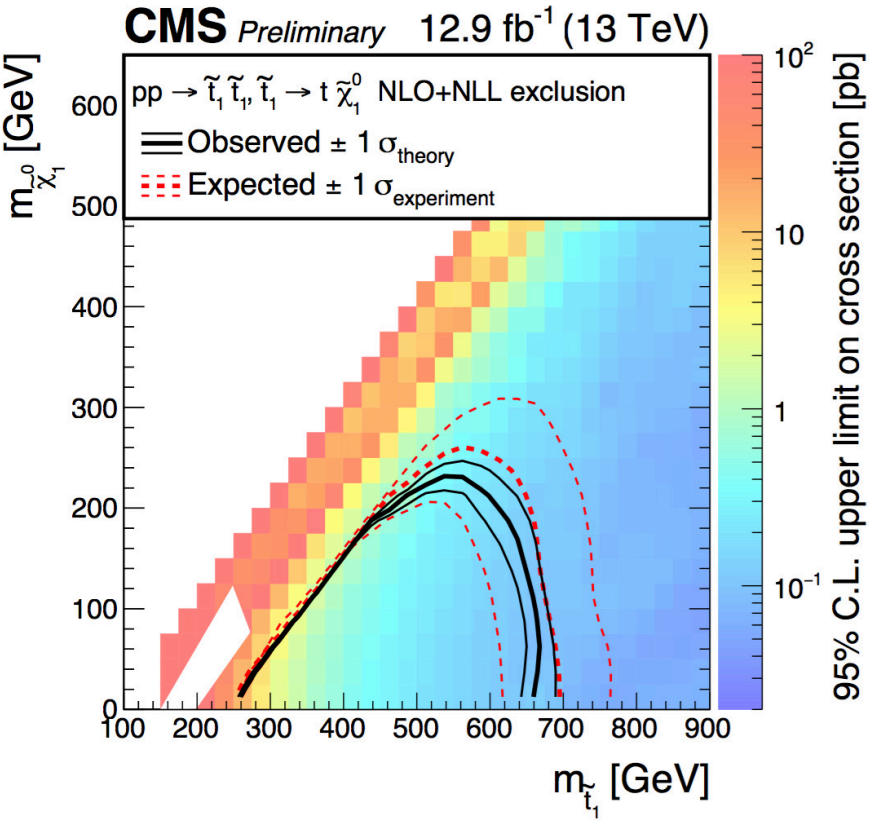


Figure 2.7: Limits on the top squark and neutralino masses in the context of a simplified SUSY model with strongly produced squarks, as derived from a search in the $t\bar{t}\tilde{\chi}_1^0\tilde{\chi}_1^0$ final state. The z axis gives the model-independent 95% C.L. upper limit on the $pp \to \tilde{t}\tilde{t}^* \to t\bar{t}\tilde{\chi}_1^0\tilde{\chi}_1^0$ cross section for the given $\tilde{t}$ and $\tilde{\chi}_1^0$ masses, determined purely by the sensitivity of the search to a final state with the corresponding kinematics. The phase space outside the observed exclusion limit (solid black line) is excluded in the scenario that the cross section for the $\tilde{t}$ production and decay is that predicted by the simplified SUSY model. The expected limit describes the region of sensitivity to the simplified SUSY model, as determined by comparing the expected signals from the SM-only and simplified SUSY scenarios.

CHAPTER 3

DETECTOR

3.1 LHC

3.1.1 Layout

The Large Hadron Collider (LHC) was first proposed in 1984 [28] by an international collaboration to investigate physics at the TeV scale. The LHC was intended as a discovery machine, searching primarily for evidence of the Higgs boson and additionally for signs of physics beyond the SM. To ensure the maximum discovery reach, the LHC was designed as a proton-proton (pp) circular collider capable of operating at center-of-mass energies up to 14 TeV. Collisions between hadrons are capable of probing a range of energy scales, due to the varying fractions of the hadron momentum carried by the partons participating in the hard interaction. Proton-proton colliders are further able to achieve high beam intensities due to the relative abundance of protons, in contrast to proton-anti-proton colliders such as the Fermilab Tevatron. The 2012 discovery of the Higgs boson [29] was a milestone for the LHC, proving its capability to probe previously unobserved physics at the TeV scale. Currently, the LHC continues to conduct precision tests of the SM while searching for evidence of new physics.

The LHC was constructed at CERN, utilizing the preexisting tunnel from the Large Electron-Positron (LEP) collider. The tunnel is 26.7 km in circumference, passing under the Franco-Swiss border at a depth of 45 to 170 m. A diagram

of the LHC complex is shown in Figure 3.1. A chain of circular accelerators of increasing size gradually accelerates protons to the target energy. These circular accelerators exploit the relatively low proton synchrotron radiation to provide significant acceleration over multiple circuits of the ring. Protons are generated from an ion source, and accelerated to 50 MeV through the Linac2 linear accelerator. The protons pass through the booster, which accelerates them to 1.4 GeV, and the Proton Synchrotron, which accelerates them to 25 GeV. The Super Proton Synchrotron then accelerates the protons to 450 GeV before injecting them into the LHC, which accelerates the protons to the target energy, up to 7 TeV.

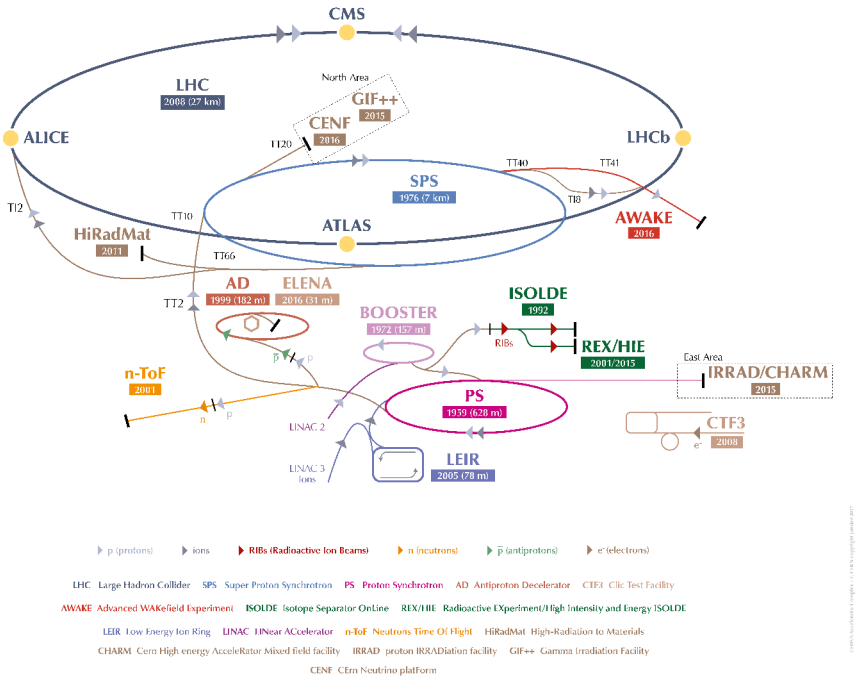


Figure 3.1: CERN accelerator complex, including LHC [30].

The following discussion of the LHC follows the presentation given in [31, 32]. The LHC is composed of eight arcs and eight straight segments. Each straight segment is approximately 528 m long. The 1st, 2nd, 5th, and 8th straight

segments contain the ATLAS, ALICE, CMS, and LHCb experimental installations, respectively. The 3rd and 7th straight segments contain beam cleaning installations, while the 6th segment contains the beam dump mechanism. The beam dump mechanism consists of kicker magnets used to deflect the beam outside the LHC ring, followed by magnets to guide the beam to a downstream target capable of absorbing the particle fluence and thermodynamic energy of the beam. The 4th straight segment contains the RF cavities responsible for accelerating the beam. Two sets of eight RF cavities, with a potential of 2 MV per cavity, accelerate protons by 485 keV / turn. The RF cavities operate at a maximum frequency of 400 MHz, driven by individual klystrons. To accommodate the RF frequency, the proton beam is divided into bunches separated by an integral multiple of the RF period. In the 2016 data taking period, this bunch spacing was 25 ns.

The eight LHC arcs contain a combination of dipole, quadrupole, and corrector magnets. The dipole magnets are responsible for bending the beam around the ring. Each arc contains 138 superconducting dipoles which generate a field of up to 8.33 T, corresponding to a bending radius of 2.8 km. The dipoles have a two bore configuration, containing both proton beams within the same magnet. The dipoles are operated at 1.9 K using superfluid helium. The dipoles must be protected against quenches, which occur when the magnet ceases to be superconducting through heating up or developing a short. The current through the quenched magnet must be dissipated quickly in order to prevent physical damage. The quadrupole magnets, which focus the beam in one plane while defocusing in the other, are used to contain beam spread. Alternating quadrupoles which focus or defocus in a given plane create a magnet matrix which contains the beam in the transverse direction. Transverse oscillations x about the ideal

trajectory s due to the quadrupole matrix may be parameterized as

$$x(s) = A\sqrt{\beta_x(s)}\cos\left(\phi(s) + \phi_0\right) \tag{3.1}$$

where A and ϕ_0 are constants which depend on the initial conditions, and $\beta(s)$ and $\phi(s)$ depend on the focusing strength along the trajectory. In the transverse plane, these oscillations around the beam direction lead to an elliptical beam profile. The beam profile is described by the emittance ϵ, where $\pi\epsilon$ is the transverse beam area.

3.1.2 Performance

The goal of the LHC is to deliver high collision intensities. Collisions occur when bunches of protons are crossed through each other in designated interaction regions, with multiple pp interactions per bunch crossing. Collision intensity is parameterized through the luminosity $\mathcal{L}$, which relates the cross section σ of a given process to the number of times it occurs via $N = \mathcal{L}\sigma$. The luminosity is a function of the beam parameters, and can be expressed as

$$\mathcal{L} = \frac{N^2 k_b f \gamma}{4\pi\epsilon_n \beta^*} F \tag{3.2}$$

where N is the number of protons per bunch, k_b is the number of bunches per beam, f is the revolution frequency, γ is the relativistic gamma factor, ϵ_n is the normalized transverse beam emittance, β^* is the beta function $\beta(s)$ at the interaction point, and F is the reduction factor due to the crossing angle. F may

be parameterized as

$$F = (1 + (\frac{\theta_c \sigma_z}{2\sigma^*})^2)^{-1/2} \tag{3.3}$$

where θ_c is the crossing angle between the two beams, σ_z is the RMS bunch length, and σ^* is the transverse RMS beam size. During the 2016 run, the LHC targeted instantaneous luminosities on the order of $1.0 \times 10^{34} cm^{-2} s^{-1}$, reaching a peak instantaneous luminosity at CMS of $1.53 \times 10^{34} cm^{-2} s^{-1}$ [33]. The beam parameters used to attain this instantaneous luminosity are given in Table 3.1. The cumulative luminosity delivered by the LHC to CMS in 2016 was $41.07 fb^{-1}$ [33].

Increasing luminosity increases the rate of interesting physics processes, but it also increases the rate of soft pp interactions known as pileup. The high inelastic cross section for pp interactions leads to a high pileup rate in the LHC, with each bunch crossing containing on average 27 pileup events at CMS in 2016. Since pileup interactions are the result of soft collisions, the particles produced are generally low-energy and aligned with the beam axis. Particle detectors must be able to properly filter and account for this low-energy background in order to accurately identify and reconstruct the results of hard interactions. Pileup mitigation is one of the factors influencing the design of the LHC detectors.

Energy (TeV)	6.5
Bunch spacing (ns)	25
β^* (m)	0.4
Crossing angle (μrad)	185×2
Number of colliding bunches	2076
Protons per bunch	1.18×10^{11}
Emittance (μm)	2.6
Bunch length (ns)	1.05
Peak Luminosity ($cm^{-2}s^{-1}$)	1.1×10^{34}
Peak mean pileup	39
Average mean pileup	27

Table 3.1: LHC beam parameters at CMS for the 2016 LHC run [34]. The beam energy, bunch spacing, number of bunches, protons per bunch, and bunch length are global beam parameters, while the remaining parameters are specific to the CMS interaction region.

3.2 CMS

The Compact Muon Solenoid (CMS) detector is one of two high-luminosity, general-purpose detectors at the LHC. A detailed description of CMS may be found in [35], which the following summarizes. To remain sensitive to a wide range of collision output, CMS aims to provide accurate identification and momentum measurement of photons, electrons, muons, jets, and missing transverse momentum over a wide kinematic range. Specifically, CMS aims to provide muon, diphoton, and dielectron mass resolutions of 1% at 100 GeV, and unambiguous muon charge identification below 1 TeV. To remain functional in a high-luminosity environment, the detector must be able to withstand high rates of particle flux. The distinguishing features of CMS, as indicated by the name, are a 3.8 T superconducting solenoid, a compact homogenous electromagnetic crystal calorimeter, and an outer layer of muon chambers to improve muon detection. CMS is 21.6 m long and 14.6 m in diameter, weighing 12500 tons. It is located in the 5th LHC straight section, below Cessy, France.

A cutaway diagram of the CMS detector is shown in Figure 3.2. Closest to the interaction point, the tracker measures the trajectories of charged particles. An electromagnetic calorimeter measures the energy of charged particles, while a hadronic calorimeter measures the energy of hadron jets. The tracker and calorimeters are surrounded by a superconducting solenoid, which provides a magnetic field of 3.8 T within the tracking volume. The charge and momentum of charged particles can be determined from the curvature of their trajectories in the magnetic field. Finally, muon chambers are located outside of the magnet. These allow for better identification and momentum measurement of muons, whose high mass and long lifetime cause them to travel through the entire detector without stopping or decaying. A depiction of interactions between final state particles and the CMS detector is shown in Figure 3.3. Information collected by the detector is read out through a trigger system. Details on these components are given in the sections below.

The CMS coordinate system is defined such that the x axis points towards the center of the LHC, while the y axis points upward vertically. The z axis then points along the beamline, towards the right when facing the center of the LHC. The azimuthal coordinate ϕ gives the angular distance in the x-y plane from the x axis, while the coordinate θ gives the angular distance from the z axis in the y-z plane. Instead of θ, angular distances in the y-z plane are usually described via the pseudorapidity $\eta = -\ln(\tan(\theta/2))$. The radial coordinate r gives the distance from the z axis in the transverse x-y plane. When protons collide, the partons involved in the hard interaction carry an unknown fraction of the proton momentum. This results in the collision rest frame receiving a significant, unknown boost along the z axis. The kinematics of the collision in the transverse plane are thus more sensitive to the physics of the hard colli-

sion. In particular, more energetic collisions will produce particles with higher transverse momenta (p_T). In addition, since the pp collision is head-on, the particles produced by the collision are expected to have zero net momentum in the transverse plane. Non-interacting particles such as neutrinos can therefore be identified from p_T imbalances known as missing transverse momentum. Because of this, the CMS detector is optimized to have the best resolution in the transverse plane.

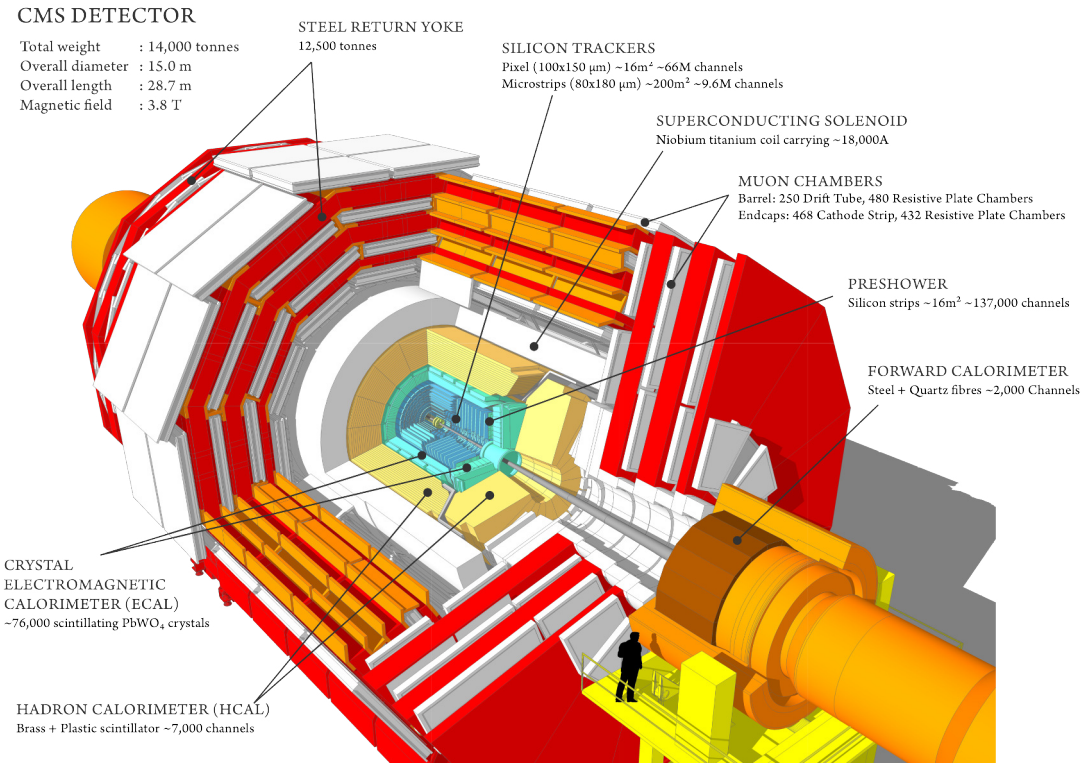


Figure 3.2: Cutaway view of the CMS detector, including overview of subsystem details [36].

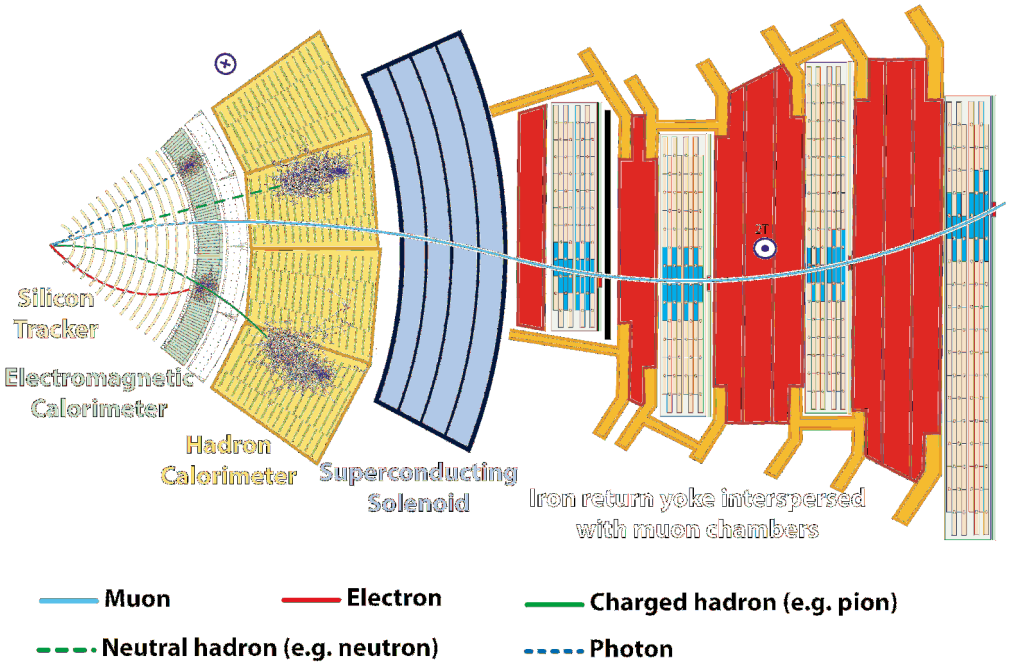


Figure 3.3: Transverse CMS detector wedge, showing interactions between final state particles and the detector [37].

3.2.1 Tracker

The CMS tracker is 5.8 m long and 2.5 m in diameter, covering $|\eta| < 2.5$. It consists of two components, an inner silicon pixel tracker and an outer silicon strip tracker. Charged particles passing through the tracker interact with the silicon, creating 3D tracks. As the tracking volume is located within the 3.8 T magnetic field, the particle charge and momentum may be determined from the track curvature. Primary and secondary vertices are identified from the intersection of particle tracks, allowing determination of the primary interaction point and subsequent particle decays. At a luminosity of $10^{34} cm^{-2} s^{-1}$, approximately 1000 particles will traverse the tracker every 25 ns. To provide accurate tracking in this environment, the tracker must be fast, high-granularity, and radiation hard. These three requirements motivate the choice of a purely silicon-based detector.

A diagram of the tracker is shown in Figure 3.4. The silicon pixel tracker, described in detail in Chapter 4, has the highest granularity and lies closest to the interaction point. The silicon strip tracker lies outside of the pixel tracker. Four components make up the strip tracker. The tracker inner barrel (TIB) and tracker inner disks (TID) form the inner part of the strip tracker, while the tracker outer barrel (TOB) and tracker endcaps (TEC) form the outer part. The TIB contains four layers at radii between 20 and 55 cm, while the TID contains three disks per endcap. Both the TIB and TID use silicon strips of length 10 cm and thickness 320 μm, with widths of 80-120 μm in the TIB and 100-141 μm in the TID. Strips in the inner layers have an occupancy of 2-3%, corresponding to the probability of a hit occuring in a given strip for a particular bunch crossing. The TOB is 236 cm long, containing six layers with radii betwen 55 and 116 cm. Strips in the TOB are 500 μm thick, with widths of 122-183 μm. The TEC contains nine disks per endcap, spanning the range $124 < |z| < 282$ cm. The disks have a 22.5 cm inner radius and 113.5 cm outer radius, and contain seven rings of radial strips. The strips are 320-500 μm thick and 97-184 μm wide. Strips in the outer layers may be up to 25 cm long. The occupancy in the TOB and TEC is ~1%. The first two layers of the TIB and TOB, as well as the first two rings of the TID and the first, second, and fifth rings of the TEC, have a second layer of strips mounted back-to-back at a stereo angle of 100 mrad. The second layer of strips provides a measurement of the z coordinate in the barrel, and the r coordinate in the disks. The overall position resolution in the TIB is 23-35 μm in ϕ and 230 μm in z. The TOB provides a position resolution of 35-53 μm in ϕ and 530 μm in z. Combined, the strip tracker yields a momentum resolution of 1-2% for particles at 100 GeV in $|\eta| < 1.6$. The track impact parameter, which describes the minimum separation between the track and the primary vertex, is used to

distinguish between particles produced in the primary interaction and particles produced from decays of unstable particles. The tracker provides an impact parameter resolution of 30-100 μm in z for 100 GeV tracks. The entire tracker is 0.4-1.8 radiation lengths thick.

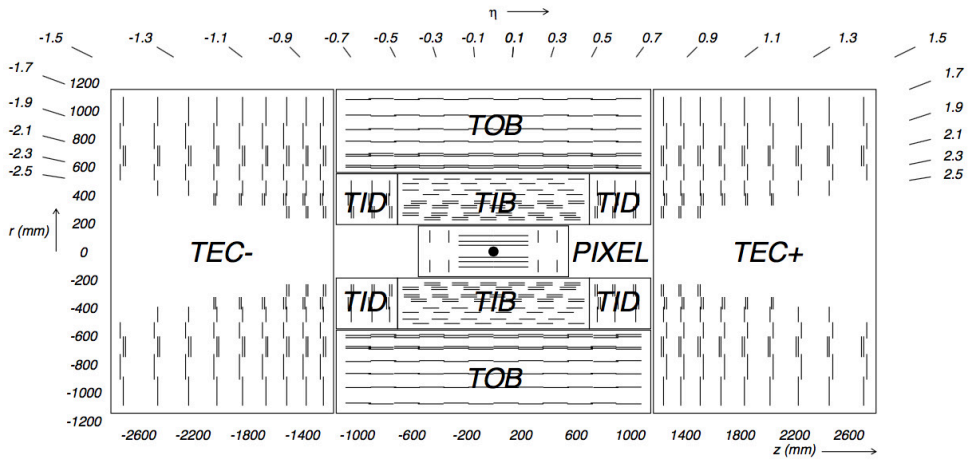


Figure 3.4: Cross section of the CMS tracker, showing the pixel and strip tracker components [35]. Lines represent detector modules.

The silicon strip tracker uses p-on-n silicon sensors, with p+ strips implanted into n type bulk and a uniform n+ implantation on the back plane connected to positive voltage. The ratio of implant width to strip pitch is 0.25, and the sensor depletion voltages are between 100 and 300 V. The strip signals are AC coupled to the readout electronics, shielding the readout from high leakage currents after irradiation. To avoid thermal effects from leakage currents and prevent damage from reverse annealing, the strip tracker is operated at $-10\,^{\circ}$C. Sensors are rectangular in the barrel and wedge-shaped in the endcap, containing either 512 or 768 strips. The strip tracker contains 24244 sensors and 9.3M strips in total, yielding a total active area of 198 m^2.

3.2.2 Electromagnetic Calorimeter

The electromagnetic calorimeter (ECAL) is designed to measure electromagnetic energy with high precision, yielding an energy resolution of ~1% at 20 GeV and ~0.4% at 100 GeV. To this end, the ECAL is constructed as a homogenous lead tungstate ($PbWO_4$) scintillating crystal calorimeter covering up to $|\eta| < 3.0$. Lead tungstate is high density, with a radiation length of 0.89 cm and a Molière radius of 2.2 cm. This permits a compact, high granularity calorimeter. Lead tungstate is also a fast scintillator, with 80% of scintillation light emitted in the 25 ns LHC bunch crossing length. The degree of scintillation is temperature dependent, requiring a steady operating temperature. An operating temperature of 18 ± 0.05 °C is achieved through water cooling.

The ECAL is divided into the barrel (EB) and endcap (EE) components. The EB covers $|\eta| < 1.479$, with an inner radius of 1.29 m. The EB contains 61200 crystals, giving a 360-fold granularity in ϕ and 170-fold granularity in η. The crystals have a 22 mm^2 front face and 26 mm^2 back face, and are 230 mm (or 25.8 radiation lengths) long. To compensate for the 0.35-0.5 mm crystal spacing, crystals are misaligned by 3% with respect to the nominal interaction vertex so that the gaps do not align with the interaction point. The EE covers $1.479 < |\eta| < 3.0$, starting at $|z| = 315.4$ cm. Each endcap consists of two half-disks, each of which contains 3662 crystals. Endcap crystals have a 28.62 mm^2 front face and 30 mm^2 back face, and are 220 mm (or 24.7 radiation lengths) long. Endcap crystals are misaligned by 2-8° with respect to the nominal interaction vertex. A diagram of the ECAL layout is shown in Figure 3.5.

Scintillation light is collected by photodetectors attached to the back face of the scintillation crystals. The low light output of lead tungstate (4.5 photoelec-

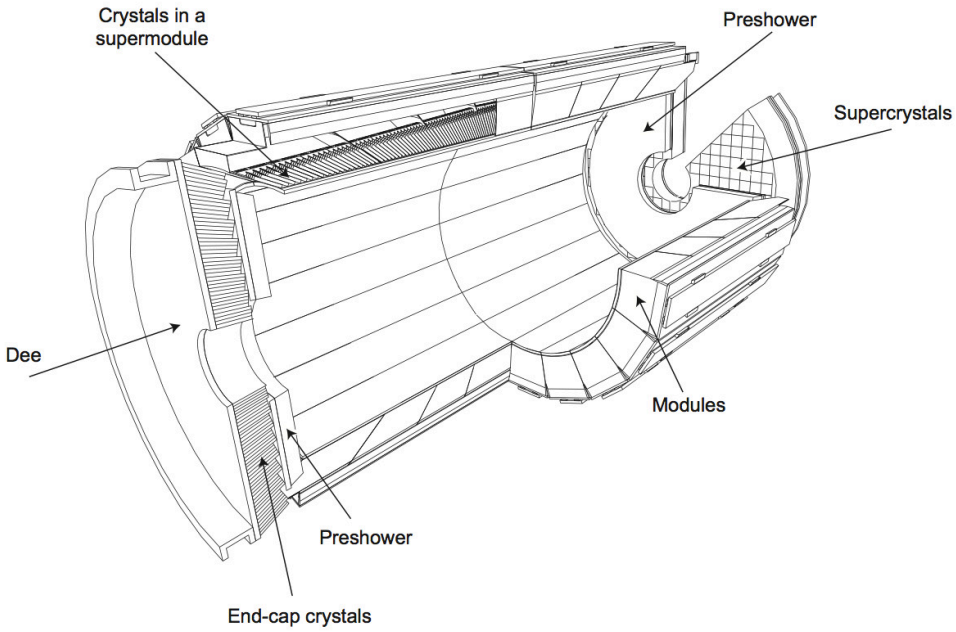


Figure 3.5: Layout of the ECAL, showing both EE and EB components [35].

trons / MeV) requires the photodetectors to operate with high amplification, in addition to being fast and radiation hard. In the EB, each crystal is connected to a pair of avalanche photodiodes. The avalanche photodiodes operate at a bias voltage of 340-430 V and produce a gain of 50. The bias voltage stability must be within tens of mV. In the EE, each crystal is connected to a vacuum phototriode. The phototriode consists of a cathode connected to ground, a dynode at 600 V, and an anode at 800 V. The phototriode has a single gain stage with a total gain of 10. The anode sensitivity varies by ~25% across the endcap, with the more sensitive anodes placed at higher radii to yield roughly constant sensitivity to p_T.

A preshower detector located in front of the EM improves pion rejection and position resolution of photons and electrons. The 20 cm (or 3 radiation lengths) thick preshower alternates two layers of lead radiator with two layers of silicon

strip sensors. EM showers are initiated in the lead, then sampled by the silicon. The silicon sensors have an active area of 61×61 mm^2 divided into 32 strips, with strips in the two layers oriented orthagonally.

3.2.3 Hadronic Calorimeter

The HCAL contributes to the measurement of hadron jets and missing transverse momentum, with a transverse energy resolution of ~0.2-0.5% at 20 GeV. The HCAL is split into four components. The hadron barrel (HB) and hadron endcap (HE) components lie within the solenoid coil, while the hadron outer (HO) and hadron forward (HF) components lie without. In total, the HCAL covers $|\eta| < 5.2$. A diagram of the HCAL is shown in Figure 3.6.

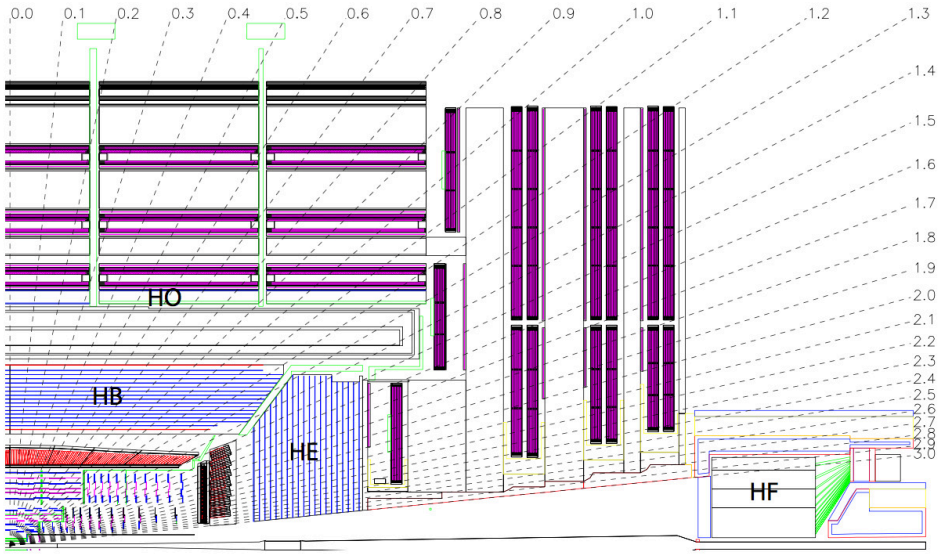


Figure 3.6: Cross section showing the various components of one quarter of the CMS HCAL [35].

The HB is a sampling calorimeter composed of brass absorber plates inter-

leaved with plastic scintillator, covering $|\eta| < 1.3$. It is divided along z into two half-barrels, with 36 ϕ wedges per half-barrel and 4 ϕ sectors per wedge. The absorber plates are arranged so that there are no radial dead areas. The scintillator is further divided into 16 η sectors per half cylinder, giving a $(\Delta\phi, \Delta\eta)$ granularity of (0.087, 0.087). The absorber plates in each sector comprise a layer of 40 mm steel, eight layers of 50.5 mm brass, six layers of 56.6 mm brass, and a layer of 75 mm steel, with the steel adding structural stability. The scintillator is 3.7 mm Kuraray SCSN81 plastic, embedded with 0.94 mm wavelength shifting fibers (WLS). The WLS are spliced to clear fibers, which are conected to a hybrid photodiode, composed of a photocathode at -8 kV separated by 3.3 mm from a silicon photodiode. The hybrid photodiode provides a gain of 2000, with $< 2\%$ resolution in the light yield. A different plastic scintillator layer is placed in front of the first absorber layer in order to detect hadronic showers due to material in front of the HCAL. The entire HB is 5.4 - 10.3 interaction lengths long. The HE, covering the range $1.3 < |\eta| < 3.0$, is constructed similarly to the HB. 17 layers of 79 mm brass absorber are interleaved with the same plastic scintillator as used in the HB. The HE has a $(\Delta\phi, \Delta\eta)$ granularity of (0.087, 0.087) for $|\eta| < 1.6$ and (0.17, 0.17) for $|\eta| > 1.6$. The ECAL and HE together are 10 interaction lengths long.

The HO extends the HB outside of the solenoid volume, improving the hadron shower containment for $|\eta| < 1.3$ and helping to identify late starting showers. The HO is composed of plastic scintillator attached to the inside of the solenoid return yoke, using the iron of the yoke as an absorber. The solenoid return yoke is divided into five rings in z. The outer four rings have a single layer of HO scintillator at a radius of 4.07 m, while the inner ring has two layers of scintillator at radii of 3.82 and 4.07 m, separated by a 19.5 cm thick iron absorber.

Each ring is divided into 12 wedges in ϕ, separated by 75 mm steel beams. The scintillator is 10 mm Bicron BC408, embedded with the same WLS as used in the HB/HE. The HO has a $(\Delta\phi, \Delta\eta)$ granularity of roughly (0.087, 0.087). The HO and solenoid extend the calorimeter depth to 11.8 interaction lengths in the central region.

The HF is located 11.2 m from the interaction point, covering the range $3.0 < |\eta| < 5.0$. The high particle flux in this region, reaching ~10 MGy for 500 fb^{-1}, requires a particularly radiation hard calorimeter. The HF is composed of quartz fibers oriented parallel to the beam line embedded in a 165 cm (10 interaction lengths) steel absorber. Particles above the Cherenkov threshold, ~190 keV for electrons, generate Cherenkov light when interacting with the steel; a small portion of this light is trapped in the quartz fiber. The light is then transmitted to shielded photomultiplier tubes. Fibers are spaced 5 mm apart in a square grid. Half of the fibers run the full length of the absorber, while the other half terminate 22 cm from the side of the absorber facing the interaction point. This allows differentiation between EM showers, which deposit most of their energy in the front of the absorber, and hadronic showers. Fibers are read out in bundles, giving the HF a $(\Delta\phi, \Delta\eta)$ granularity of (0.175, 0.175).

3.2.4 Superconducting Solenoid

In CMS, the momentum of a charged particle is determined by measuring the curvature of the particle track in a known magnetic field. In order to maintain good resolution at high momentum, a strong magnetic field is needed. A strong magnetic field also improves position resolution in the tracker and calorime-

ters by spreading energy and charge deposits across multiple cells. A uniform, constant magnetic field of 3.8 T is generated in the tracking and calorimeter volumes by a superconducting solenoid. The solenoid is 12.5 m long, with a 6 m inner diameter and 6.3 m outer diameter. The thinness of the solenoid necessitates a self-supporting coil. NbTi reinforced with aluminum alloy provides the necessary structural stability. The winding is composed of four layers with a total of 41.7×10^6 ampere-turns. The stored energy in the solenoid is 2.6 GJ. Current flows in the superconductor with no dissipation at 6.44 K, though the solenoid is operated at 4.5 K as a safety margin. Superfluid helium is used as a coolant, due to its high thermal conductivity and low viscosity. The magnet cold mass is 220 tons. The magnet flux is returned through a 13 m long, 14 m outer diameter iron yoke. The yoke weighs 10000 tons and is divided into five wheels and two endcaps, with three disks per endcap.

3.2.5 Muon Chambers

Detecting and precisely measuring the momentum of muons is a main goal of CMS. To this end, a series of muon detectors forms the outer CMS layer. The muon chambers yield a 95-99% muon reconstruction efficiency in regions with no gaps and a 9% standalone momentum resolution at 200 GeV. When tracker information is included, the momentum resolution improves by a factor of 10. The muon chambers comprise drift tubes (DTs) in the barrel and cathode strip chambers (CSCs) in the endcap. Resistive plate chambers (RPCs) offer an independent muon trigger, as well as improved position resolution.

The DTs cover $|\eta| < 1.2$. Individual cells are 1.99 - 3.02 m long, and 13×42 mm

in cross section. A gold-plated stainless steel anode wire runs down the middle of the cell and is maintained at 3.6 kV. 16 mm aluminum tape field electrodes run along the wide sides of each cell, while 11.5 mm aluminum tape cathodes run along the narrower sides. The field electrodes and cathodes are maintained at 1.6 and -1.2 kV, respectively. The cell is filled with a 85% Ar, 15% CO_2 gas mixture with a gas gain of 10^5. As muons pass through the chamber, the gas is ionized and charge collects on the anode wire to be read out. The time delay of the signal readout gives a position coordinate along the wire, with the 265 ps clock resolution leading to a 250 μm position resolution along the wire length. The average cell occupancy is 0.76%. The DT assembly is composed of five rings, corresponding to the five rings of the magnet return yoke. Each ring is divided into 12 chambers in ϕ and 4 stations in r. The first three stations contain three superlayers per chamber, while the outermost station contains only two superlayers per chamber. Each superlayer contains fours layers of tubes staggered by a half-cell. The first two superlayers in each chamber are oriented such that the wires run parallel to the beam axis, giving a coordinate in the r-ϕ plane. The third superlayer, where present, is oriented such that the wires run perpendicular to the beam axis and provide a coordinate in z. A diagram of the DT station placement is shown in Figure 3.7.

The CSCs cover $0.9 < |\eta| < 2.4$, where the range $0.9 < |\eta| < 1.2$ overlaps with the DTs. A diagram of the CSC placement is shown in Figure 3.8. CSCs are multiwire proportional chambers containing six anode wire planes interleaved with seven cathode planes, surrounded by a gas medium. The CSCs are designed to operate at a high rate and provide optimal performance despite non-uniform magnetic field, gas temperature, or gas pressure. CSC segments are wedge shaped, covering 10° or 20° in ϕ. In each wedge the anode wires are

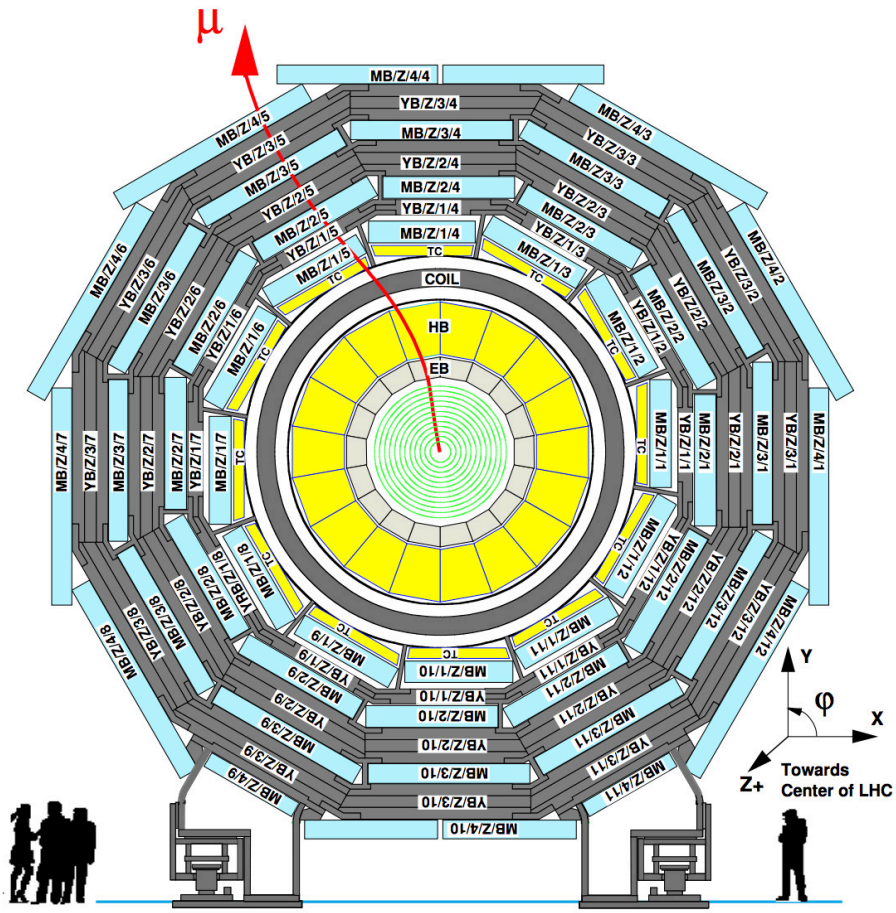


Figure 3.7: Cross section showing the placement of DT stations in relation to the magnet return yoke [35]. DT stations are shown in light blue.

oriented azimuthally and give the r coordinate of a hit, while the cathode strips are radial and give the ϕ coordinate. The cathode strips are 8.4 mm wide at the narrow end of the wedge, widening to 16 mm. The anode wires are gold-plated tungsten, spaced 3.2 mm apart with ~1000 wires per panel. The gas gap between subsequent cathode layers is 9.5 mm, containing a gas mixture of 40% Ar, 50% CO_2, and 10% CF_4. The gas gain is ~7×10^4 at an operating voltage of 3.6 kV. The CSCs have a 99% track stub efficiency and provide an offline resolution of 70-150 μm in the r-ϕ plane.

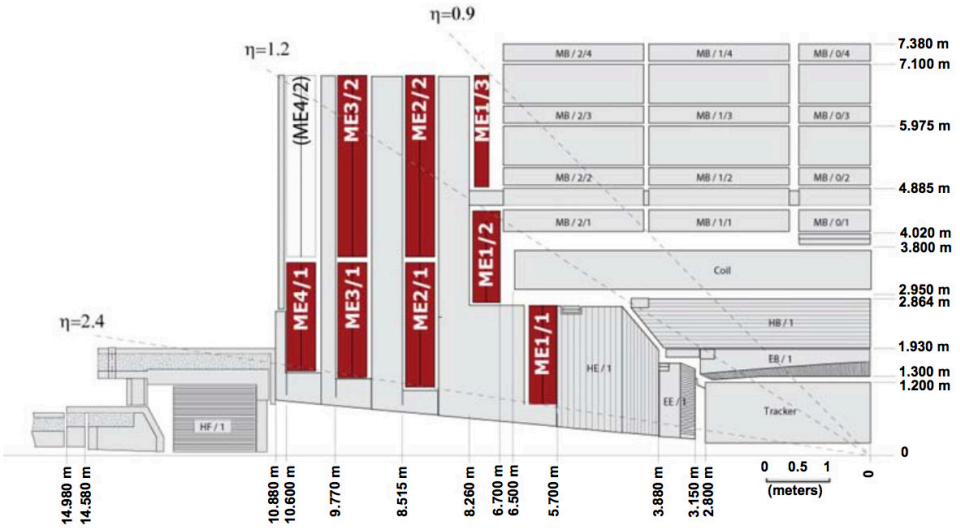


Figure 3.8: Cross section showing the placement of CSC stations in one quarter of CMS [35]. CSC stations are shown in red.

The RPCs cover $|\eta| < 1.6$ and provide a dedicated independent trigger for the muon chambers. The RPCs provide high time resolution for hits, unambiguously identifying the source bunch crossing. The RPCs are gaseous parallel-plate detectors, containing readout strips sandwiched between two gas gaps. The gas is 96.2% $C_2H_2F_4$, 3.5% iC_4H_{10}, and 0.3% SF_6, with added water vapor for 45% humidity. There are six RPC layers in the barrel, two each at the first two DT stations and one each at the last two. A diagram of the RPC placement is shown in Figure 3.9. RPC strips in the barrel run parallel to the beam axis, covering 2.455 m in z with a width of $5/16^\circ$ in ϕ. In the endcap, three layers of RPCs are located at the first three CSC stations. Each layer is composed of three rings of trapezoidal wedges containing radial strips. Wedges in the inner ring cover 20° in ϕ and wedges in the outer two rings cover 10°.

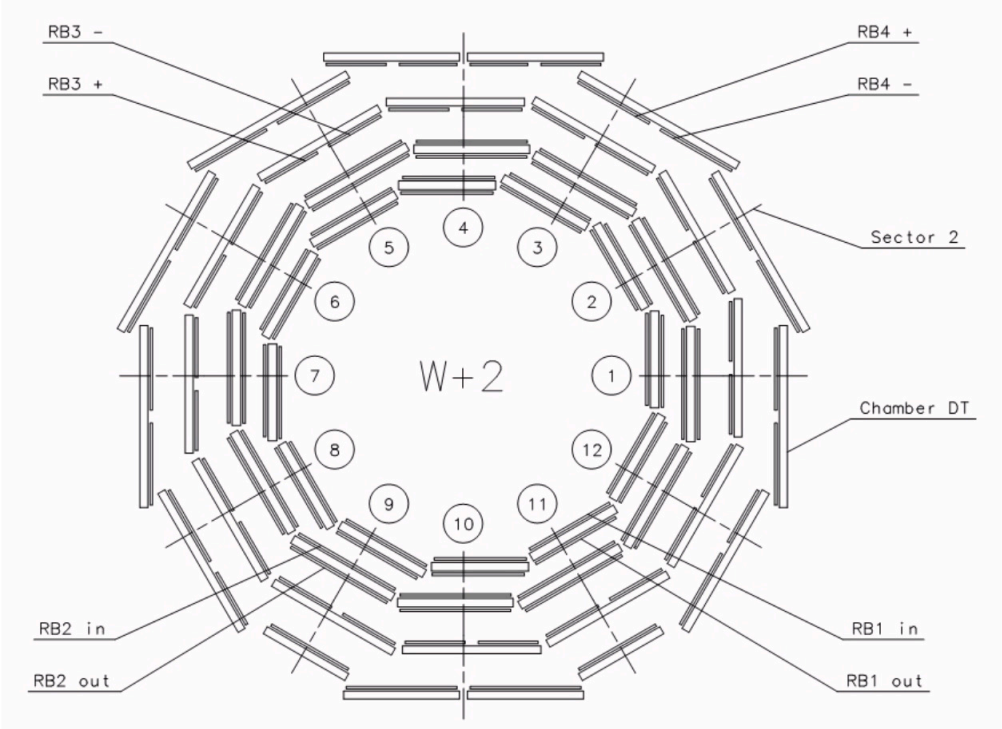


Figure 3.9: Cross section showing the placement of RPC stations in relation to DT stations in the barrel [35].

3.2.6 Trigger

At the LHC, bunches of protons collide at the rate of 40 MHz. In addition to the primary interaction, approximately 20 soft pp interactions are expected to occur in a given bunch crossing. To cope with such a high data rate, fast and efficient filtering algorithms are necessary to reduce the data flow to a rate which may be written to memory while retaining potentially interesting physics events. The CMS trigger system uses a two-step process to reduce the event readout rate from 40 MHz to 100 Hz. The level-1 (L1) trigger is implemented in programmable electronics, and relies on coarsely segmented data from the calorimeters and muon chambers to roughly identify interesting particles. The L1 trigger reduces the event rate from 40 MHz to 100 kHz. The high-level trigger

(HLT) then uses a filter farm of over a thousand processors to make decisions using the full detector data, reducing the event rate from 100 kHz to 100 Hz. The events which pass the HLT are then written to memory for offline analysis.

L1 Trigger

At the most basic level, the inputs to the L1 trigger are trigger primitives generated by the calorimeter and muon chambers. Regional triggers identify the most interesting objects in a segment of the calorimeter or muon chambers, which are sent to a global calorimeter or muon trigger. The global calorimeter and muon triggers then select the most interesting objects in the calorimeter and muon chambers to send to the global L1 trigger, which decides whether to accept or reject the event.

In the calorimeter, trigger primitives are calorimeter towers, which correspond to 5×5 ECAL crystal clusters and corresponding HCAL deposits. The regional calorimeter trigger identifies which trigger primitives most resemble photons and electrons, based on shower shape information and energy deposit ratios between the HCAL and ECAL. The regional calorimeter trigger also identifies the transverse energy sum in the given region. The global calorimeter trigger further identifies HCAL jets and determines the overall missing transverse energy.

The muon trigger recieves input from the DTs, CSCs, and RPCs. DT trigger primitives are track segments, consisting of matched track stubs from multiple superlayers. CSC trigger primitives are 3D track segments using information from the anode and cathode readout. RPC trigger primitives identify spatially

and temporally coincident hit patterns. A regional trigger covering 30° in the barrel or 60° in the endcap matches track segments between different DT or CSC stations. Finally, the global muon trigger matches DT and CSC tracks to RPC hits and interpolates isolation information from the global calorimter trigger. The most interesting objects from the muon and calorimeter triggers are sent to the L1 global trigger, which decides whether to pass the event on to the HLT. The L1 trigger process takes 3.2 μs.

HLT

The HLT is a more dynamic filter than the L1 trigger, and may be modified to reflect changes in luminosity or physics goals. More details on the HLT are given in [38]. In most cases, the inputs to the HLT are L1 trigger objects, known as seeds. Information from the parts of the detector associated with the L1 seed is used to perform a more thorough object reconstruction and ID. To minimize CPU usage, events are discarded after each reconstruction step if they fail to meet quality criteria. HLT trigger objects include electrons, muons, taus, photons, jets, and p_T^{miss}. To respect the needs of different offline analyses, various HLT working points are defined for each object. Typically, each working point includes an ID requirement and restrictions on the object p_T or η. An object might have a low p_T and tight ID, or loose ID and high p_T, or medium requirements on both aspects. HLT triggers require one or more HLT objects. Typically, multi-object triggers will relax requirements on individual objects. Similar triggers are grouped into trigger streams, which are assigned a given portion of the HLT bandwidth. In some cases, a trigger may select interesting events but operate at too high a rate. In this case, the trigger is prescaled such that only a

given fraction of the events passing the trigger are saved. The triggers assigned to each trigger stream and their prescales constitute a trigger menu. Trigger menus are designed using input from all offline analysis groups, to ensure that each group receives an appropriate fraction of the total dataset. Different trigger menus are defined for different levels of luminosity, in order to maintain appropriate overall HLT rates.

3.2.7 Data Handling

The CMS data acquisition (DAQ) system coordinates operation and readout of the CMS subdetectors. More information on the DAQ can be found in [38]. To cope with the dataflow required for the HLT, the DAQ system must be able to read out a data flow of 100 GByte/s coming from approximately 650 data sources. The DAQ system is modular by subdetector. Each subdetector stores data in 40 MHz buffers, to be read out upon receipt of a L1 trigger. Event fragments from the subdetectors are sent to the DAQ, where an event builder assembles the event fragments into a complete event. An event filter runs the HLT on the full event and writes passing events to memory. These events are sampled for data quality monitoring purposes. The run control and monitoring system (RCMS) provides a web interface to the DAQ, allowing the user to view the detector state and perform state transitions (such as configuring the detector or starting data-taking). The detector control system manages all detector electronics, providing monitoring and control over power supplies as well as safety and temperature monitoring.

Offline analysis of CMS data is handled through a worldwide computing

network. The CMS software [39] package provides a common platform for event reconstruction and data analysis. CMS data is stored in several formats, of increasing order of processing and reconstruction. The RAW data format contains the full detector output after passing the HLT selection. The RECO data format applies particle ID, track and vertex reconstruction, and filtering algorithms to compress the event size while generating high-level physics objects. The AOD (analysis object data) format filters and compresses only the high-level physics objects from the RECO data, producing light datasets with sufficient information for most analysis needs. The MINIAOD format operates similarly to AOD, but at a higher level of compression.

The CMS computing network is arranged in a tier structure. The Tier 0 is located at CERN and handles permanent storage of RAW data, initial reconstruction of RAW data to RECO, and the export of RAW and RECO data to Tier 1 centers. Tier 1 centers provide a backup storage site for RAW data, as well as storage for RECO, AOD, and MINIAOD datasets. Tier 1 centers also provide computing resources to analyze large datasets. Tier 2 centers provide computing resources to analyze smaller datasets, as well as production of simulated data. While Tier 1 centers are managed by national labs and require continuous professional computing support, Tier 2 centers may be mangaged by collaborating universities. CMS users have access to the entire grid when performing data analysis and storage.

CHAPTER 4

PIXEL DETECTOR

This chapter gives an overview of the CMS pixel tracker. The pixel tracker is located closest to the interaction point, providing high-resolution tracking and vertexing of charged particles. The boosted $t\bar{t}$ cross section measurement described in this thesis relies on the identification of wide jets formed from the decay of a boosted top quark. These t jets are identified from jet substructure, requiring high precision in the momentum and direction of the constituent tracks. The information provided by the pixel tracker is therefore crucial for the identification of boosted top quarks.

This chapter is structured as follows. Section 4.1 gives a summary of the pixel tracker layout, following information given in [35]. Section 4.2 describes the software used to operate the detector, and Section 4.3 describes the detector configuration parameters and their storage and modification methods. Section 4.4 describes the methods used to calibrate the detector. Finally, Sections 4.5 and 4.6 describe error recovery procedures and data quality monitoring, respectively. Since my graduate studies have involved time spent operating and calibrating the pixel tracker, particular emphasis is given to these aspects.

4.1 Detector Layout

Discussion of the pixel tracker layout summarizes information given in [35, 40]. The pixel tracker is divided into a barrel component (BPix) and a forward component (FPix), with a total coverage up to $|\eta| < 2.5$. BPix is 53 cm long, and

has three layers located at radii of 4.4, 7.3, and 10.2 cm. FPix has two disks per endcap, at 34.5 and 46.5 cm in $|z|$. Both BPix and FPix are divided into four half-cylinders or disks. FPix contains 12 blades per half-disk, with each blade containing 21 read out chips (ROCs) on the side facing the interaction point and 24 ROCs on the opposite side. Each BPix half-cylinder contains eight sectors in ϕ, with each sector further divided into one to three ladders in ϕ and each ladder containing eight modules in z. Each module contains 16 ROCs, save for half-modules along the half-cylinder edges which contain 8 ROCs. A diagram of the pixel tracker layout is shown in 4.1.

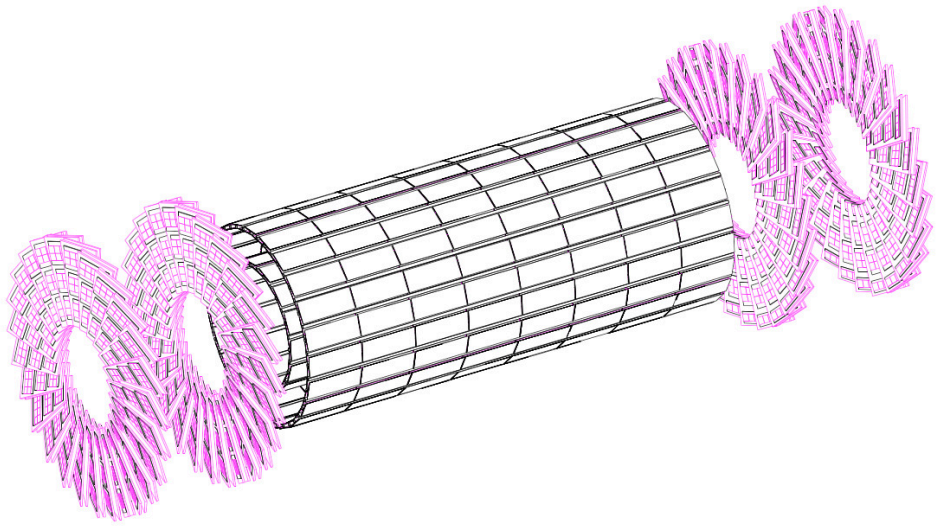


Figure 4.1: Layout of the pixel tracker, showing FPix (pink) and BPix (black) [40].

The pixel tracker sensors are n-in-n silicon, with high-dose n (n+) pixel implants in a high-resistance n-type bulk. Concentric open p-stop rings surround the n+ implants. On the opposite side of the sensor are located p implants surrounded by guard rings. This design offers high gain for low bias voltage, and operates even when partially depleted due to high radiation dose. The sensor is 180 μm thick, divided into 100×150 μm pixels with an average occupancy

of 0.01%. Charge sharing between pixels, enhanced by Lorentz drift due to the magnetic field, improves the position resolution to ~15-20 μm. The pixel p-implants are bump-bonded to ROCs, with each ROC reading out 52×80 pixels. The ROC stores a mask bit and four trim bits per pixel. The mask bit determines whether a pixel is included in the readout, while the trim bits allow per-pixel variations of the threshold which a pixel hit must exceed to be read out. Each ROC also contains a number of digital analog converters (DACs). These programmable values describe, among other things, the amount and timing of charge to inject for calibration purposes, the base threshold for all pixels on the ROC, and biases and gains which affect different aspects of the ROC output. Pixel hits which exceed the combined ROC and pixel trim threshold are transmitted by the ROC as an analog pulse height, along with analog values encoding the pixel row and column. Six analog levels are used to encode digital information, with a series of five such levels describing the pixel address. These analog levels are used elsewhere in the signal to encode other digital information. ROCs are read out in series by a token bit manager (TBM). Each ROC sends a placeholder pattern known as a ROC header, followed by any pixel hits. Each TBM handles readout for one module or half-module, corresponding to 8 or 16 ROCs in BPix and 21 or 24 ROCs in FPix. The ROC output is sandwiched between a TBM header and trailer, which encode the event number and error information, respectively. An example of the analog signal transmitted by a TBM is shown in Figure 4.2. A portcard collects the signals from six TBMs, with an analog opto-hybrid (AOH) converting the analog signal to an optical pulse which is transmitted to the front end driver (FED). The FED is a VME module with 36 input optical channels. 32 FEDs serve BPix, while 8 FEDs serve FPix. The FED digitizes the optical input, sampling the signal at a rate of 40 MHz.

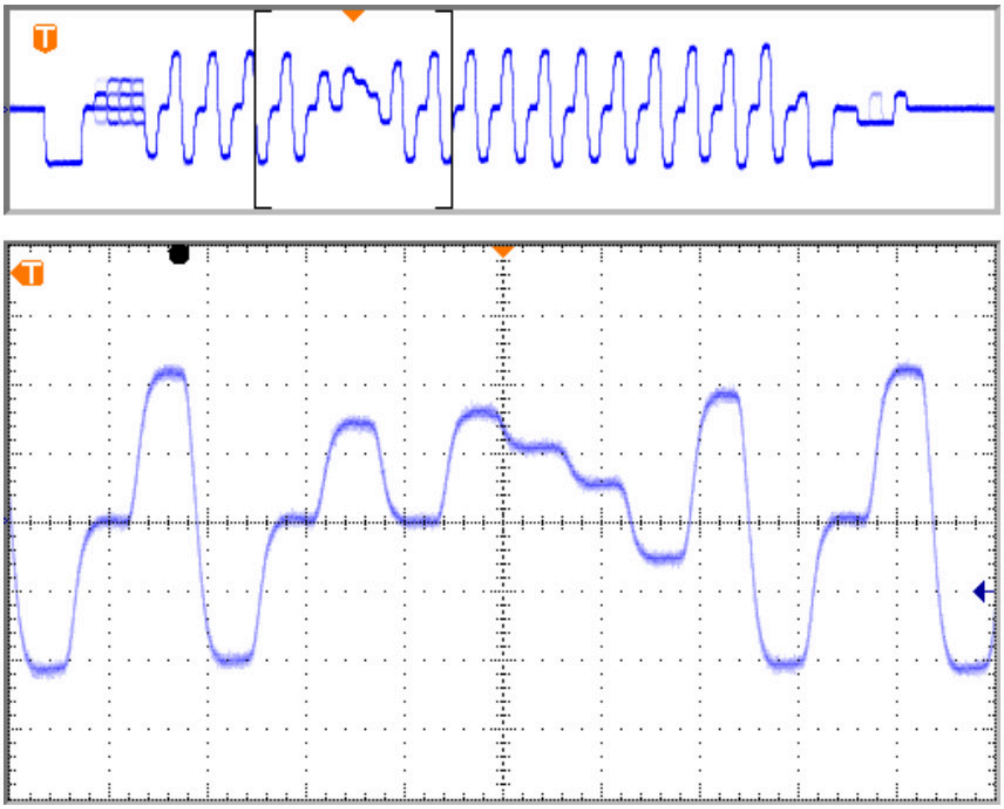


Figure 4.2: Analog signal sent by a TBM containing 16 ROCs, of which one ROC has a pixel hit [41]. The signal begins with a TBM header, followed by four ROC headers with no hits. The fifth ROC contains a hit, and the header is followed by five levels encoding the pixel row and column, then the hit pulse height. Eleven more ROC headers and a TBM trailer follow. The zoomed-in section focuses on the information sent regarding the pixel hit. The signal period is 25 ns.

The pixel tracker is controlled by the pixel front end controller (FEC). The pixel FEC transmits fast timing signals and slow control data as 40 MHz optical signals. A digital opto-hybrid (DOH) converts the optical signals to a 40 Mbit/s signal to be sent to the forward electronics. The timing signal is used to generate the clock and L1 trigger, while the control data is used to program the electronics. The data link is bi-directional, allowing data to be transmitted back to the FEC. The FEC takes 14 s to configure the detector, with the majority of the time being used to configure the pixel trim and mask bits. In addition to the pixel

FEC, a tracker FEC is used to send slow configuration commands to the AOH, DOH, and other supply electronics. A cartoon of the pixel control and readout is shown in Figure 4.3.

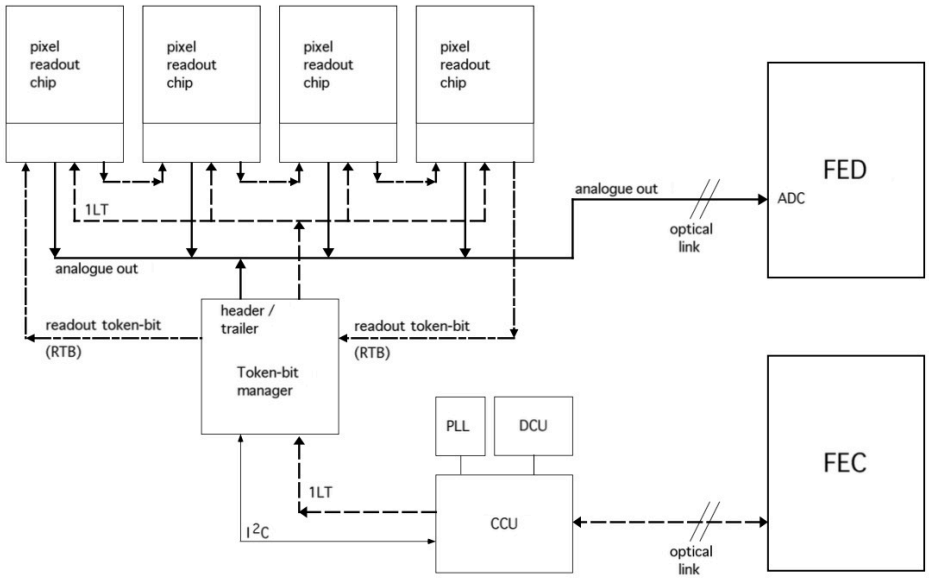


Figure 4.3: Pixel tracker readout and control [40]. The clock and trigger (single dashed line, labeled 'L1T') are sent by the FEC to the readout electronics. A token bit sent by the TBM instructs the ROCs to send data, which is collected and sent to the FED (solid line).

CAEN power supplies are used to power the pixel tracker. Each ROC is powered by two low voltage power supplies, a 1.5 V analog line and a 2.5 V digital line. A HV power supply capable of providing up to 600 V provides the silicon bias voltage. Power supply channels for BPix feed 192 ROCs, while FPix power supply channels feed 135 ROCs. The average current draw is 4.6 A on the analog line and 9 A on the digital line. Auxiliary 2.5 V lines power the AOH, DOH, and service electronics. The pixel tracker is operated at $-10\,^{\circ}\mathrm{C}$ to avoid thermal effects from leakage currents and prevent damage from reverse annealing, using C_6F_{14} as a coolant. The detector control system monitors temperatures, currents, and voltages across the pixel tracker. Values which fall

outside of normal ranges trigger notifications to relevant experts, while values which fall outside of safety tolerances cause the pixel tracker to be automatically powered off.

4.2 Pixel Online Software

The pixel online software (POS) handles the pixel tracker DAQ. The top-level application is the PixelSupervisor, which manages a number of independent supervisors controlling different elements of the DAQ hardware. These include the PixelFECSupervisor, PixelFEDSupervisor, PixelTKFECSupervisor, PixelTTCSupervisor, and PixelLTCSupervisor, which control FED crates, pixel FEC crates, tracker FEC crates, global timing (TTC), and local triggers, respectively. Communication between the supervisors is handled using the SOAP protocol. The level-1 function manager (L1FM) provides an interface between RCMS and the PixelSupervisor, propagating state change requests from RCMS to the PixelSupervisor. For local running, a GUI interface to the L1FM allows state transitions to be requested manually. Separate POS packages implement the various supervisors. These packages depend on additional packages describing the FED and FEC interfaces, configuration formats and the configuration database interface, calibration formats, common utilities, and detector components.

4.3 Configuration

The pixel tracker configuration data may be stored locally, or online in an Oracle database. POS classes describe the configuration format and provide an interface to the database. Pixel configurations are modular, comprising multiple objects describing different aspects of the configuration. Different versions of a configuration object have different values of the configuration parameters. A configuration key describes a particular set of object versions. The objects included in the configuration are listed below.

- **detconfig**: list of all ROCs in the detector, with ROCs which cannot communicate flagged as NOANALOGSIGNAL and ROCs which cannot be configured flagged as NOINIT
- **nametransation**: mapping of ROCs to FED and FEC channels
- **fedcard**: FED settings, including analog values used to encode digital information
- **dac**: ROC DAC settings
- **mask**: pixel mask bits
- **trim**: pixel trim bits
- **tbm**: TBM settings
- **portcard**: portcard settings, including AOH bias and gain and relative timing of sent and received data
- **portcardmap**: mapping of AOH channels to FED channels
- **fedconfig**: list of FEDs used
- **fecconfig**: list of FECs used

- **ttcciconfig**: list of TTCCis used
- **tkfecconfig**: list of TKFECs used

4.4 Calibrations

The pixel tracker is calibrated to ensure optimal performance. Calibrations generally produce new versions of configuration objects, with a new configuration key generated to include the new versions. Detector calibrations are divided into two general stages. An initial series of calibrations ensures that the analog signal transmitted by the TBMs has the correct shape to be intelligible when digitized by the FED. More involved calibrations then improve the stability and efficiency of the detector response to a wide range of signals. Typically, the full calibration procedure is implemented only after major alterations to the detector, such as the replacement of hardware components or change of operating temperature. However, select calibrations are performed on a more regular basis to correct for slight shifts in operating temperature or radiation damage effects.

4.4.1 Signal Digitization

The calibrations described below ensure that the analog signal can be converted into an intelligible digital signal.

Clock Phase

The first step in ensuring proper signal digitization is to set the clock phase. The FED is driven by the LHC clock, sampling the analog signal every 25 ns. The FED has a delay with a granularity of $^{25}/_{16}$ ns controlling when data is sent, and a phase controlling when data is latched. Adjusting these parameters adjusts the point at which the analog signal is sampled. The clock phase calibration scans over the 16 delay and 2 phase values, looking for the point where the FED samples the analog signal closest to the black-ultrablack transition (defined below) in the TBM trailer. This transition point provides a clear marker of the start of a 25 ns period in the analog signal. The calibration creates a new version of the fedcard object with the optimized delay and phase.

Delay25

The portcard receives clock and trigger signals from the FEC and sends pixel hit data to the FED. For successful communication, the timing between the data and the clock lines has to be correct. This timing is managed by the Delay25 chip. The Delay25 calibration scans the send data (SDa) and return data (RDa) delays to determine optimial settings for communication. For each combination of (SDa, RDa), the calibration sends a series of commands and checks the result. The scan produces an efficiency plot in (SDa, RDa) space, and the optimal settings for RDa and SDa are chosen to lie in the middle of the efficient region. The calibration creates a new version of the portcard object with the optimized values of SDa and RDa.

FED Baseline

The FED baseline calibration adjusts the input offset and channel offsets of the optical receivers in the FED such that the signal baseline is set to 450 ADC units, near the midpoint of the FED dynamic range. The baseline is the neutral value of the analog signal, and is sent when no information is being conveyed. It is also referred to as the black level and is the second analog level used to encode digital information. The first, lower, analog level is referred to as the ultrablack (UB) level. In addition to setting the baseline as close to 450 ADC units as possible, the calibration also saves the actual analog values of the black and UB levels. The calibration creates a new version of the fedcard object with optimized settings. This calibration must be run frequently, as the FED baseline is highly temperature dependent. The calibration is also useful for diagnostic purposes, producing a plot of the digitized analog signal. As no pixel hits are being read out during the calibration, the analog signal is expected to contain a TBM header, 8-24 ROC headers, and a TBM trailer. All headers contain a black and UB level, which are expected to be consistent. A ragged-looking signal, with inconsistent blacks or UBs, may indicate a poorly calibrated or bad channel. A flat signal may also occur if there are problems communicating with a given channel. An example of a digitized signal for a well-calibrated channel is shown in Figure 4.4.

AOH Bias

The AOH bias setting controls the amount of light sent from the portcard to the FED for a given channel. The AOH bias setting affects the black and UB levels differently, and is chosen to optimize the black-UB separation. Initially,

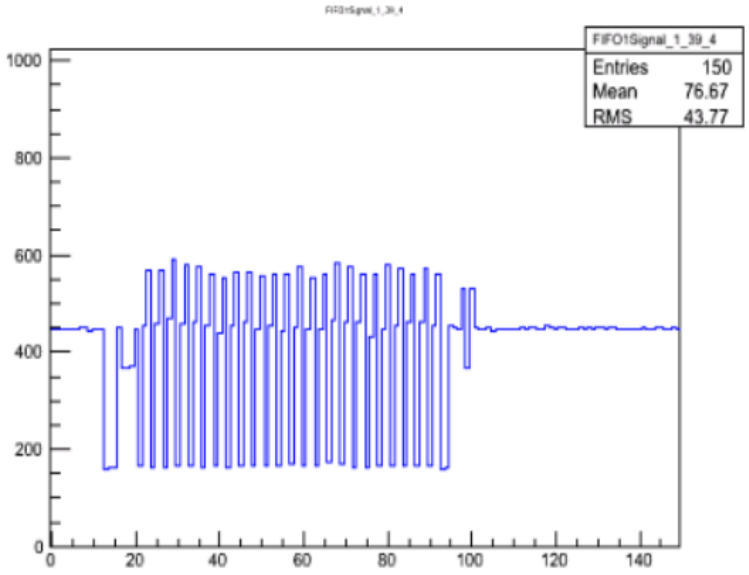


Figure 4.4: Digitized signal from a single TBM, produced as output of the FED baseline calibration. This channel is well calibrated, with consistent TBM and ROC UBs ($\approx$ 175 ADC units, on y axis) and blacks (450 ADC units).

as the AOH bias increases there is no effect on either the black or UB levels of the signal. At some point, the black level begins increasing linearly with the AOH bias while the UB level remains constant. Later, the UB level begins to increase, with the same dependence on the AOH bias. The optimal value of the AOH bias is right after the UB level begins to increase, where the black-UB difference is saturated but the AOH is not drawing unnecessary current. After the AOH bias has been optimized, the calibration performs a coarse adjustment of the FED baseline, to bring the signal back into the dynamic range of the FED. The AOH bias calibration produces new versions of the portcard and fedcard objects, with optimized AOH bias and FED baseline, respectively.

AOH Gain

The AOH gain scales the size of deviations from the black level. Each channel has four possible settings of the AOH gain. The AOH gain calibration attempts to set the AOH gain high enough that the UB in the TBM header is below a configurable value, assuming that the black level is at 450 ADC units. The AOH gain is not set any higher to minimize power consumption. The AOH gain calibration produces a new version of the portcard with optimized AOH gain settings.

TBM UB

The TBM UB calibration performs a finer adjustment of the TBM UB levels, targeting a specific UB level for all TBMs. The TBM contains three DACs affecting the TBM UB, known as AnalogInputBias, AnalogOutputBias, and AnalogOutputGain. Values for these DACs are scanned simultaneously. The TBM UB calibration produces a new version of the tbm object with updated settings for the AnalogInputBias, AnalogOutputBias, and AnalogOutputGain.

ROC UB

The ROC UB calibration sets the ROC UB levels equal to the TBM UB level. Each ROC contains two DACs affecting the ROC UB, known as VIbias_roc and VIbias_DAC. Both affect the levels used to encode digital information, but only VIbias_roc affects the analog pulse height. Therefore, VIbias_roc is held fixed and VIbias_DAC scanned. The ROC UB calibration produces a new version of the dac object, with updated values of VIbias_DAC.

Address Levels

The address levels calibration determines the six analog levels used to encode digital information. These levels are known as address levels since they are typically used to encode the row and column of pixel hits. Since the actual values of the levels have been set by previous calibrations, this calibration merely checks the consistency and separation between individual levels. It also stores the values of each level for use by the FED. The calibration generates multiple analog signals encoding digital information, such that each level is replicated multiple times. Figure 4.5 shows an example of multiple such signals overlaid on top of one another. The analog values for all levels are read out, and the mean and variance for each level determined. The calibration checks that six address levels exist, where the level separation is significantly larger than the level variance. The calibration then produces a new version of the fedcard object, containing the values of the six address levels.

Pixel Alive

The pixel alive calibration is a diagnostic, checking the readout efficiency for each pixel on a ROC. For each pixel, a series of hits above the ROC threshold are sent and readout attempted, yielding a hit efficiency. Maps of the hit efficiency for all pixels on a ROC are produced. Regions of low or zero efficiency may be due to physical damage or poor configuration settings. In particular, inefficient ROCs are often caused by the ROC threshold being set too low. As the pixel alive calibration is diagnostic, it does not produce any new configuration data.

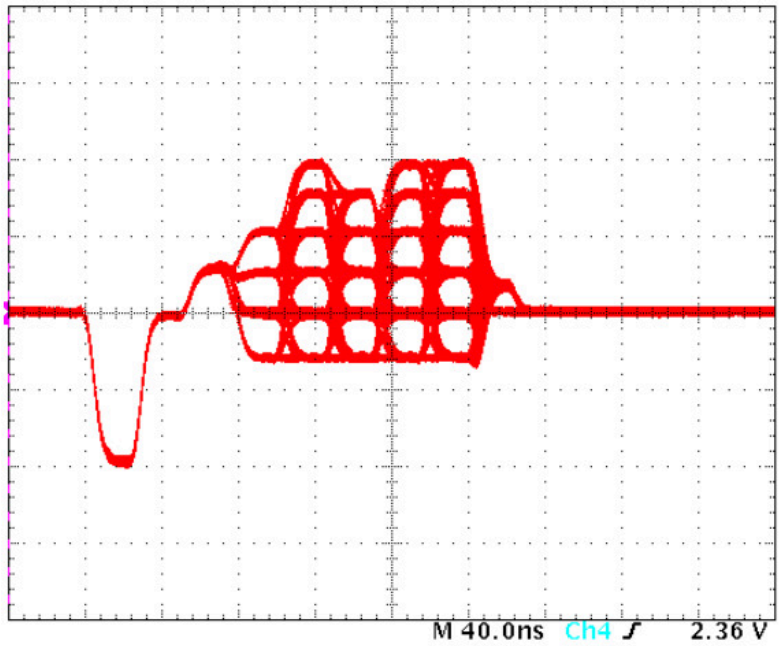


Figure 4.5: Overlay of multiple analog signals corresponding to different digital patterns, showing the consistency of the analog levels used to encode digital information [41]. The six analog levels along the y axis are the six address levels. Patterns of five such levels, as are overlaid here, are used to encode pixel row and column information. The levels shown are consistent and well-separated, allowing for dependable conversion of analog signal to digital information.

4.4.2 Signal Quality

The calibrations described ensure the detector response remains stable and efficient with respect to a wide range of signals.

Vana

The Vana calibration adjusts the ROC pre-amplifier bias current (Vana) to optimize signal rise speed while remaining within power supply limits. The analog pulse is gaussian in shape, and is typically spread across several 25 ns bunch crossings. Increasing Vana decreases the pulse width, but Vana cannot be in-

creased beyond ~24 mA per ROC without exceeding the power supply limit. The signal speed can be assessed by comparing the in-time and absolute thresholds, where the in-time threshold describes the smallest signal to cross the ROC threshold in the triggered bunch crossing and the absolute threshold describes the smallest signal to cross the ROC threshold in any bunch crossing. DT, the difference between the in-time and absolute thresholds, decreases with increasing signal speed and reaches zero when the signal peaks within the triggered bunch crossing. The Vana calibration adjusts DT for all ROCs to a uniform target value, such that the average current draw per ROC is ~24 mA.

The Vana calibration is iterative. In each step, the in-time and absolute theshholds are calculated for each ROC and Vana adjusted to bring DT towards the target value. The thresholds are determined by taking SCurve scans. In an SCurve scan, pixels are injected with varying amounts of calibration charge and readout efficiencies measured for each value. The result is an S-shaped turn on curve, where the efficiency rises rapidly from 0 to 1 when the amount of charge injected passes the threshold. The Vana calibration typically takes ten iterations, or roughly two hours, to complete.

Threshold Minimization

The threshold minimization calibration adjusts the ROC thresholds to be as low as possible, without becoming sensitive to noise. This allows the detector to be sensitive to the maximum dynamic range of input signals. The ROC threshold is controlled by the VcThr DAC, but the relation between the value of VcThr and the actual threshold varies with radiation damage. Because of this, the threshold minimization procedure must be performed periodically to update

VcThr. When the threshold is too low and the readout is sensitive to noise, pixel hits cannot be read out. ROCs which are sensitive to noise will produce a high number of inefficient pixels in the Pixel Alive calibration.

The threshold minimization procedure is iterative. The threshold is initially set to a high value, such that the readout is known to be insensitive to noise. A Pixel Alive calibration is taken to verify that the ROCs are still efficient. The threshold is then lowered step by step, with subsequent Pixel Alive calibrations taken at each step. The ROC is considered inefficient when the number of inefficient pixels rises above a given cutoff, indicating that the ROC is sensitive to noise. Inefficient ROCs are recovered by raising the threshold back to the last good value, plus a small safety margin. Once the ROC recovers, no further adjustments to the threshold are made. The calibration continues until all ROCs have become inefficient and been recovered. A SCurve scan is then taken to measure the actual thresholds and look for any outlier ROCs. Outlier ROCs may then have parameters adjusted by hand.

Pulse Height Optimization

The pulse height optimization procedure ensures that the dynamic range of the analog signal amplitude is as large as possible, while maintaining a linear relation between the amount of charge deposited in the pixel and the corresponding signal amplitude. The pulse height optimization procedure is a series of four calibrations. The Vsf and VHldDel calibrations adjust the ROC DACs which affect the linearity between the deposited charge and signal amplitude. The PHRange calibration then maximizes the dynamic range of the signal, while remaining within the limits of the FED dynamic range. Finally, the gain calibra-

tion measures the exact relation between the amount of deposited charge and the analog signal amplitude. The Vsf DAC controls the linearity of the pixel response, with higher values giving stronger linearity but also increasing current draw. ROC testing prior to assembling the detector revealed that Vsf values which give equal pulse heights for low and high values of the DAC VHldDel also give good linearity. In the Vsf calibration, distributions of the pulse height as a function of Vsf are produced for low and high values of VHldDel. The intersection point of these curves gives the optimal Vsf. Once Vsf is optimized, the VHldDel calibration determines the value of VHldDel which maximizes the pulse height. The ROC contains four additional DACs which affect the pulse height: VIbias_PH, VOffsetRO, VOffsetOp, and VIon. The PHRange calibration fixes VOffsetRO and VIon and scans over VIbias_PH and VOffsetOp to determine the settings which give the maximum pulse height while remaining within the FED dynamic range. Finally, the gain calibration measures the exact relation between charge deposited in the pixel and the pulse height by injecting varying amounts of charge and measuring the resultant pulse heights. The output of the gain calibration is used for offline reconstruction.

Determining Trim Bits

The pixel trim bits, scaled by the ROC DAC Vtrim, provide an additional correction to the per pixel threshold. The pixel trimming procedure adjusts the trim bits to yield a uniform pixel threshold across the ROC. To determine the trim bits, the per pixel thresholds are measured for varying values of Vtrim and the trim bits, where the threshold measurements are taken from SCurve scans. An interpolation is performed to determine the optimal values of Vtrim and

the trim bits. To save time, Vtrim is determined using a subset of pixels on the ROC, after which the trim bits are determined for all pixels. Typically, the pixel trimming is only performed once for a given ROC.

Final Masking

The pixel mask bits are used to remove pixels from the readout which send signals in the absence of actual deposited charge. These hot pixels are generally caused by radiation damage, requiring periodic updates of the masking. Hot pixels are identified through measuring the pixel occupancy when the only real source of charge deposition is from cosmic particles. Pixels with a high occupancy, typically $> 0.1\%$, are masked.

4.5 Error Recovery

After calibration, the detector is in an optimal state for data-taking. However, radiation damage, temperature fluctuations, and other effects can cause the data transmitted by the pixel tracker to become corrupted. The most common cause of corrupted data is a single event upset (SEU). This occurs when a highly ionizing particle causes a bit flip in some part of the pixel memory. If the SEU occurs in one of the buffers used to store readout data, the datastream becomes corrupted. In this case, the data is cleared and the event containing the corrupted data is skipped. If the SEU affects the stored pixel trim and mask bits, the resultant improper calibration will cause the ROC to send bad data. Channels containing ROCs with this type of SEU are removed from the readout until the ROC can be reconfigured with the correct trim and mask bits. In addition to

SEUs, radiation damage to the pixel sensors and temperature fluctuations can alter the optimum configuration of the detector. If temperature fluctuations or radiation damage cause a channel to begin sending bad data, the channel is removed from the readout until it can be recalibrated.

4.6 DQM

The pixel data quality monitoring (DQM) provides fast feedback on the quality of the data being sent from the pixel tracker. The DQM provides a web GUI to view quantities such as the number, charge, and size of pixel clusters; the pixel occupancy and number of dead ROCs; and summaries of the errors sent by the FED in processing data. DQM quantities are provided as a function of the CMS run, allowing characterization of time-dependent effects. The DQM is divided into online DQM and offline DQM, with the online DQM providing fast monitoring as data is collected and the offline DQM utilizing the full offline reconstruction.

CHAPTER 5

EVENT RECONSTRUCTION

This thesis presents a measurement of the $t\bar{t}$ cross section for very energetic top quarks, using events where one top quark decays hadronically via $t \to Wb \to q\bar{q}b$ and the other decays leptonically via $t \to Wb \to \ell\nu b$. The lepton ℓ may be a muon or electron. At high energies, the jets produced by the hadronization of the $q\bar{q}b$ final state quarks merge into a single wide jet known as a t jet. This measurement thus attempts to identify events containing a t jet, a b jet, a muon or electron, and missing transverse momentum p_T^{miss} from the non-interacting neutrino. This chapter describes the reconstruction of these final state objects from CMS data.

The reconstruction begins by constructing inner charged particle tracks [42], calorimeter clusters [43], and muon tracks [44, 45] by associating hits in the tracker or deposits in the calorimeter. The Particle Flow (PF) algorithm [43] then uses these elements to reconstruct all stable final state particles, as described in detail in Section 5.1. The electrons, muons, photons, charged hadrons, and neutral hadrons reconstructed by the PF algorithm are then used to construct higher-level objects. First, charged hadrons that are not associated with the primary vertex are classified as pileup candidates. These are ignored in the subsequent object reconstruction through the charged hadron subtraction (CHS) scheme [46]. PF electrons and muons which pass further quality criteria are considered final state electrons and muons. These requirements are described in Sections 5.2 and 5.3, respectively. Next, jet clustering is performed on all PF particles that are not classified as pileup candidates. The jet clustering does not exclude the PF electrons and muons which correspond to final state electrons

and muons. These PF electrons and muons are removed from the jets by hand to avoid double-counting. The jet energies are corrected to account for variations in the detector response. More details on the jet reconstruction may be found in Section 5.4. The p_T^{miss} is calculated from the PF particles, accounting for jet corrections and cleaning, as described in Section 5.5. Finally, tagging algorithms are used to identify jets produced by a bottom or top quark. The b tagging and t tagging algorithms are described in more detail in Sections 5.6 and 5.7, respectively.

5.1 Particle Flow

The PF algorithm aims to reconstruct the type, direction, and energy of all stable final state particles in an event. The algorithm utilizes charged particle tracks in the inner tracker, energy deposits in the ECAL and HCAL, and muon tracks to reconstruct muons, electrons, photons, and charged and neutral hadrons. These particles are then used to construct more complex objects, such as jets, tau leptons, and p_T^{miss}. The algorithm works by linking together inner tracks, calorimeter clusters, and muon tracks which appear to be produced by the interaction of the same final state particle with the detector. Linked elements are grouped into a PF block, from which one or several final state particles is reconstructed. A detailed description of the PF algorithm is given in [43].

Inner tracks may be linked to calorimeter clusters in the ECAL, HCAL, or preshower (PS). Tracks are extrapolated into the appropriate calorimeter, and linked to any calorimeter clusters which are separated from the track by at most one calorimeter cell. Additionally, ECAL clusters are linked to a track if they match tangents extrapolated from the track to the ECAL, consistent with the

emission of a bremsstrahlung photon. Tracks compatible with conversion of a bremsstrahlung photon are also linked to the original track if the reconstructed photon trajectory is tangent to the original track. Calorimeter clusters in different detectors are linked if the position of the higher-granularity cluster (PS or ECAL) is within the envelope of the lower-granularity cluster (ECAL or HCAL). Inner tracks are linked to muon segments if the extrapolated position of the inner track in the muon system is compatible with the muon segment. Inner tracks and muon tracks are linked if the hits from the two tracks can be fit as a single global track. Inner tracks are linked to other inner tracks if the tracks share a common displaced vertex, such that the tracks are consistent with a nuclear interaction between a charged hadron and the tracker.

Linked inner tracks, calorimeter clusters, and muon tracks form a PF block. From each block, the PF algorithm attempts to reconstruct muons, electrons, charged hadrons, neutral hadrons, and photons. If a particle is reconstructed from the block, the corresponding elements are removed from the block before proceeding.

The PF algorithm first attempts to reconstruct a muon from the block, exploiting the clean muon signature. The muon may be a global muon, standalone muon, or tracker muon. Global muons are reconstructed from a linked inner track and muon track with good global fit quality. Global muons must either be isolated or pass strict track quality criteria. Standalone muons are reconstructed from a muon track, while tracker muons are reconstructed from a linked inner track and muon segment. Neither standalone muons nor tracker muons may be linked to significant calorimeter deposits. For muons with an inner track and $p_T < 200$ GeV, the muon momentum is taken from the inner track. Otherwise,

the muon momentum is taken from the best track fit.

The PF algorithm next attempts to reconstruct an electron from the block. Electrons are reconstructed from linked inner tracks and ECAL clusters, including all tracks and ECAL deposits compatible with bremsstrahlung radiation. The primary inner track is refit with a Gaussian-sum filter (GSF) [47] to account for kinks due to bremsstrahlung. Electrons must have a high GSF fit quality, good track-cluster matching, and little energy in the HCAL. The electron energy is calibrated to account for energy lost due to missed calorimeter clusters.

Finally, the algorithm attempts to reconstruct photons, charged hadrons, and neutral hadrons from the remaining inner tracks and calorimeter clusters in the block. ECAL and HCAL clusters which fall within the tracker acceptance but are not linked to any inner tracks are reconstructed as photons and neutral hadrons, respectively. Standalone ECAL clusters which fall outside the tracker acceptance are reconstructed as photons, while standalone HCAL clusters and linked HCAL and ECAL clusters which fall outside the tracker acceptance are reconstructed as hadrons. ECAL and HCAL clusters linked to inner tracks are used to simultaneously reconstruct photons, charged hadrons, and neutral hadrons. If the net calorimeter energy is compatible with the sum of track momenta, a charged hadron is reconstructed. If the calorimeter energy is instead significantly larger than the track momentum, photons or neutral hadrons may be reconstructed in addition to the charged hadron. If the HCAL energy alone exceeds the track momentum, the HCAL excess is used to reconstruct a neutral hadron and any ECAL excess is used to reconstruct a photon. Otherwise, the combined HCAL and ECAL excess is used to reconstruct a photon. Alternatively, if the track momentum is significantly larger than the calorimeter energy,

the algorithm attempts to reconstruct low-quality muons or fake tracks before reconstructing the charged hadron.

Inner tracks linked to a common secondary vertex are used to reconstruct the nuclear interaction of a charged hadron with the detector. If an incoming track exists for the vertex, the incoming track direction and momentum are used for the charged hadron. The momenta of outgoing particles are corrected to maintain conservation of momentum at the vertex.

Once all PF blocks have been reconstructed as particles, a final set of selections and corrections is applied to reduce overly high p_T^{miss} resulting from poorly reconstructed muons. Poorly reconstructed muons include cosmic muons misidentified as collision muons, muons with poorly reconstructed p_T, and punch-through charged hadrons misidentified as muons. If removing a cosmic muon, using an alternate measurement of the muon p_T, or replacing a muon and neutral hadron pair with a charged hadron reduces the p_T^{miss} by more than 50%, the change is applied.

5.2 Electrons

Signal electrons are identified as PF electrons which pass an additional set of quality criteria. The combined reconstruction and selection efficiency for signal electrons is 70-90%, depending on the electron p_T and η [48, 49]. The misidentification rate is 0.1%. The quality criteria include requirements on the size and shape of the ECAL supercluster, matching between the supercluster and inner track, fraction of calorimeter energy in the HCAL, agreement between track momentum and calorimeter energy, track quality, and track impact parameter with

respect to the primary vertex. Different requirements are used for EB electrons as opposed to EE.

In addition to these quality criteria, signal electrons are also required to be isolated. This requirement distinguishes between electrons produced in the hard interaction and electrons originating from a jet. The standardized isolation requires that all particles within a cone of $\Delta R = \sqrt{\delta\phi^2 + \delta\eta^2} = 0.3$ around the electron have a sum p_T less than a given fraction of the electron p_T. Unfortunately, the leptons produced in leptonic decays of boosted top quarks are often highly collimated with the b jet, and thus fail the standardized isolation requirement. Instead, this measurement utilizes the mini isolation algorithm [50], where the isolation cone radius scales with the electron p_T. Specifically, electrons are required to have miniIso = $p_T^{cone}/p_T^e < 0.1$. More details on the mini isolation algorithm are given in Appendix A. Finally, signal electrons are required to have $p_T > 50$ GeV and $|\eta| < 2.1$. Scale factors (SFs) are applied to compensate for differences between data and MC in modeling the electron ID and isolation efficiencies [51].

5.3 Muons

Signal muons are identified as PF muons which pass an additional set of quality criteria. The combined reconstruction and selection efficiency for signal muons is over 99%, and the misidentification rate is ~1% [52, 53]. Signal muons are required to be global or tracker muons with more than 80% valid tracker hits. Tracker muons must have good matching between the inner track and muon segment. Global muons must have a slightly looser matching between the inner

track and muon segments, a good global track fit, good matching between the inner track and muon track, and few kinks.

In addition to these quality criteria, signal muons are also required to be isolated. Similarly to electrons, mini isolation is used to yield improved performance in the highly boosted regime. Muons must have miniIso < 0.1. Signal muons are also required to have $p_T > 50$ GeV and $|\eta| < 2.1$. SFs are applied to compensate for differences in modeling the muon reconstruction, ID, and isolation efficiencies between data and MC [54].

5.4 Jets

Jets are reconstructed by clustering PF final state particles using the anti-k_t algorithm [55]. The anti-k_t algorithm is fast, producing jets which are roughly conical and whose kinematics are insensitive to collinear and soft radiation. The last property, referred to as infrared and collinear (IRC) safety, ensures that the jet properties are not affected by the soft and collinear approximations used to model parton showering. Given two particles or particle clusters i and j, the algorithm defines the distance properties

$$d_{ij} = min\left(\frac{1}{p_{Ti}^2}, \frac{1}{p_{Tj}^2}\right) * \frac{(y_i - y_j)^2 + (\phi_i - \phi_j)^2}{R^2} \tag{5.1}$$
$$d_{iB} = \frac{1}{p_{Ti}^2}$$

where R is the jet radius. The rapidity $y = \frac{1}{2}\ln\left(\frac{E+p_z}{E-p_z}\right)$ is somewhat analogous to η, though it depends on the particle energy. If $d_{ij} < d_{iB}$ the objects are

clustered, while if $d_{ij} > d_{iB}$ the object i is identified as a jet and excluded from further clustering. Clustering proceeds in order of angular separation. In a final state containing several hard particles and many soft particles, the algorithm produces conical jets of radius R around hard particles which are separated by more than 2R from the nearest hard particle, near-conical jets centered around hard particles separated by more than R but less than 2R, and a single near-conical jet with substructure centered around hard particles separated by less than R. The independence of the jet cones with respect to soft particles characterizes the anti-k_t algorithm as IRC safe. The anti-k_t algorithm is implemented by means of the FastJet package [56].

Two different jet collections are considered in this measurement. AK4 and AK8 jets refer to anti-k_T jets clustered using radii of $R = 0.4$ and $R = 0.8$, respectively. The CHS scheme is applied to exclude charged hadrons originating from pileup interactions. Jets must satisfy a loose set of quality criteria [57] requiring the jet to contain a non-zero amount of charged hadrons. AK4 jets are required to have $p_T > 50$ GeV and $|\eta| < 2.4$, while AK8 jets are required to have $p_T > 400$ GeV and $|\eta| < 2.4$. AK4 jets are b jet candidates, while AK8 jets are t jet candidates.

Since the PF muons and electrons which correspond to final state muons or electrons are initially used in the jet clustering, the momenta of these PF leptons is subtracted from any jet in which they are contained to avoid double-counting. No reclustering is performed after the jet cleaning.

Due to the nonlinearity and η dependence of the calorimeter response, the energy of reconstructed jets must be corrected to yield the true particle jet energy. First, a correction is applied to account for contributions to the jet energy

from pileup particles. MC simulated jets are then corrected to account for differences between the actual and simulated detector response. Next, the jets are corrected for η dependent variation in the detector response. Finally, the jet energies are corrected to account for shifts caused by the FastJet algorithm. Differences between the corrected jet energy resolution (JER) in data and MC are accounted for by smearing the MC JER to match data. More details on the jet energy correction (JEC) and JER methods can be found in [58]. The specific JEC factors used in this analysis were calculated from 2016 data and can be found in [59]. The data JER typically amounts to 15% at 10 GeV, 8% at 100 GeV, and 4% at 1 TeV.

5.5 Missing Transverse Momentum

Since protons are collided head on at the LHC, the particles produced in the collision are expected to have zero sum p_T. Non-interacting particles such as neutrinos will not be reconstructed by the detector, leading to an imbalance in the reconstructed net transverse momentum. Missing transverse momentum describes the magnitude of the net transverse momentum imbalance, defined as $p_T^{miss} = \left|\vec{p}_T^{\,miss}\right| = \left|-\sum \vec{p}_T\right|$. In the semileptonic $t\bar{t}$ final state, the p_T^{miss} is expected to be the p_T of the neutrino. However, p_T^{miss} also receives contributions from mismeasured particle p_T. The PF p_T^{miss} is measured using the full set of PF particles. The PF p_T^{miss} is corrected to compensate for JEC and JER, as well as lepton cleaning.

5.6 Tagging b Jets

B tagging algorithms identify jets from the decay of a b quark. These algorithms exploit the long b quark lifetime, identifying jets originating from a secondary production vertex displaced from the primary collision vertex. This measurement utilizes the Combined Secondary Vertex v2 (CSVv2) b tagging algorithm [60]. The CSVv2 algorithm employs a multivariate technique, analyzing the number and quality of secondary vertices in the jet, the secondary vertex displacement significance, the total number of tracks in the jet and fraction associated with the secondary vertex, the fraction of jet energy carried by particles associated with the secondary vertex, the invariant mass of particles associated with the secondary vertex, the degree to which tracks point away from the primary vertex, the angular separation between the jet axis and displacement axis, and the angular separation between tracks and the jet axis. The output of the CSVv2 algorithm is a discriminant between 0 and 1, where 1 corresponds to a b jet. This measurement requires that the CSVv2 discriminant be above 0.8484 for b-tagged jets, corresponding to a mistag rate of 1% [61].

Data/MC SFs are used to correct the b tagging efficiencies in simulation to match those in data. These SFs are measured as a function of the p_T and η of the b-jet candidate, with separate SFs for the identification of true b jets and misidentification of true u, d, s, or c jets. In this analysis, only one b jet candidate is considered for b tagging. The b jet candidate is the highest p_T AK4 jet in the same hemisphere as the lepton. Since the event only contains one b jet candidate, a simple event reweighting is sufficient to account for differences between MC and data b tagging efficiencies. MC events are weighted by SF_i if the b jet candidate is b-tagged and $(1 - \epsilon_i)/(1 - \epsilon_i * SF_i)$ if not, where ϵ_i is the b

tagging efficiency measured in MC. The MC b tagging efficiency is measured by determining the fraction of events in the signal sample where the b-jet candidate passes the b tagging requirement, for events which pass the preselection.

5.7 Tagging t Jets

A hadronically decaying top quark produces three jets, corresponding to the b, q, and $\bar{q}$ decay products. At high Lorentz boosts, these three jets merge into a single wide jet. Since the angular distance between the decay products of a particle of mass m and energy E is $\Delta\theta = 2m/E$, the decay products of a top quark with $p_T > 400$ GeV will be mostly contained within a cone of R = 0.8. T tagging algorithms are used to identify wide jets from top quark decays, using jet substructure.

In this measurement, t jets are identified through requirements on the jet substructure and groomed jet mass. The groomed jet mass is calculated using the modified mass drop tagger (mMDT) [62], also known as the 'soft drop' algorithm with soft threshold $z_{cut} = 0.1$, angular exponent $\beta = 0$, and jet radius R = 0.8. A detailed description of the generalized soft drop algorithm is given in [63]. In general, jet grooming algorithms attempt to remove from the jet wide-angle soft radiation resulting from initial state radiation (ISR), particles originating in the underlying event, and pileup. The result is a jet whose momentum is closer to that of the original parton. In the soft drop algorithm, the jet is reclustered using the Cambridge-Aachen method [64, 65], which provides an angular ordered pairwise clustering. The final clustering step is then undone, leaving two subjets. In the generalized soft drop algorithm, if the two subjets

have transverse momenta p_{T1} and p_{T2} and angular separation ΔR_{12} such that

$$\frac{\min(p_{T1}, p_{T2})}{p_{T1} + p_{T2}} > z_{\text{cut}} \left(\frac{\Delta R_{12}}{R_0} \right)^{\beta} \tag{5.2}$$

then the subjets are reclustered and the grooming is complete. Otherwise, the softer subjet is discarded and the process repeated on the harder subjet. In the case of mMDT, the above expression reduces to

$$\frac{\min(p_{T1}, p_{T2})}{p_{T1} + p_{T2}} > 0.1, \tag{5.3}$$

removing any dependence on the angular separation of the subjets. Using mMDT, the energy loss due to grooming is small and roughly independent of the strong couping α_S. When tagging a t jet, we require the groomed jet mass to be between 105 and 220 GeV.

Jet substructure is described by the n-subjettiness variable τ_N [66, 67], which gives a measure of how 'N-prong-like' a jet is. N-subjettiness is defined as

$$\tau_N = \frac{\sum_k p_{T,k} \min(\Delta R_{1,k}, \Delta R_{2,k}, ..., \Delta R_{N,k})}{\sum_k p_{T,k} R_0} \tag{5.4}$$

where $\Delta R_{n,k}$ is the angular separation between particle k and the nth subjet axis and R_0 is the jet radius. Jets with $\tau_N \sim 0$ have most constituents aligned with one of the N subjet axes, and therefore contain N or fewer subjets. On the other hand, jets with $\tau_N \gg 0$ contain a significant fraction of constituents which do not align with any subjet axis, implying N+1 or more subjets. Ideally, the subjet axes would be chosen to minimize τ_N; however, this is too computationally intensive

in practice. Instead, the subjet axes are identified through reclustering the jet with an exclusive k_T algorithm [68, 69] designed to produce exactly N subjets. Homogenous jets will generally have $\tau_N \sim \tau_{N-1}$ independent of N, whereas a jet with N subjets will have $\tau_N \sim 0$ and $\tau_{N-1} \gg 0$. Therefore, jets with N subjets are identified through an upper limit on τ_N/τ_{N-1}. In this measurement, t-tagged jets are required to satisfy $\tau_{32} \equiv \tau_3/\tau_2 < 0.81$.

The t tagging efficiency is assumed to perform differently in MC and data. Since the t tagging efficiency can only be measured in our signal region, the data/MC t tagging SF is extracted simultaneously with other parameters describing the measured cross section, as described in Section 6.7.3.

CHAPTER 6

MEASUREMENT OF DIFFERENTIAL $T\bar{T}$ CROSS SECTION

6.1 Analysis Overview

This thesis presents a measurement of the $t\bar{t}$ cross section as a function of top quark p_T, for boosted top quarks with $p_T > 400$ GeV. The $t\bar{t}$ pairs are produced when the LHC collides protons at a center of mass energy of 13 TeV, and the collision output is reconstructed using data collected by the CMS detector. The measurement is performed in the semileptonic final state, where one top quark decays hadronically via $t \to Wb \to q\bar{q}b$ and the other decays leptonically via $t \to Wb \to \ell\nu b$. The lepton ℓ may be a muon or electron. The measurement utilizes the full 2016 dataset, corresponding to an integrated luminosity of 35.9 fb^{-1}.

The boosted $t\bar{t}$ cross section measurement in the semileptonic final state is complimentary to resolved measurements at lower top quark p_T, as well as measurements in the dileptonic and all-hadronic final states. 43.8% of $t\bar{t}$ events are found in the semileptonic final state, compared to 45.7% in the all-hadronic final state and 10.5% in the dileptonic final state [70]. The semileptonic final state therefore benefits from both higher statistics and a clean lepton signature. As the $t\bar{t}$ final state topology depends on the top quark p_T, separate measurements in the low and high p_T regimes improve the measurement sensitivity. The three quarks produced in the hadronic top quark decay hadronize, producing jets in the final state whose angular separation is inversely proportional to the top quark energy. At low top quark energies, the three jets are well separated and may be individually reconstructed from CMS data. However, at high energies

the three jets merge to form a single wide t jet. Measurements optimized for low energies, which require a higher number of jets in the final state, are therefore insensitive to $t\bar{t}$ decays at higher energies. For example, a recent measurement by CMS of the differential $t\bar{t}$ cross section at 13 TeV in the semileptonic final state [71] required events to contain at least four jets. As such, the measurement was insensitive to top quark energies above ~600 GeV. The measurement presented in this thesis is instead optimized for the high energy, or boosted, regime. Events are required to contain a single wide t jet, a b jet, a muon or electron, and missing transverse momentum (p_T^{miss}) from the escaping neutrino, matching the boosted semileptonic final state topology.

Measurements of the boosted differential $t\bar{t}$ cross section in the semileptonic final state were performed at 8 TeV by both CMS [72] and ATLAS [73]. These were the first measurements of the $t\bar{t}$ cross section in the boosted topology. The boosted differential $t\bar{t}$ cross section has also been measured at 13 TeV by CMS in the all-hadronic final state [74], and by ATLAS in the semileptonic final state [75].

This chapter is structured as follows. Section 6.2 gives an overview of the data and simulation samples used in the analysis, while Section 6.3 describes the selection used to identify the final state. Section 6.4 describes the backgrounds to the $t\bar{t}$ signal, as well as the methods used to model these backgrounds. Section 6.5 shows a comparison between the observed data and prediction from simulation. Section 6.6 describes the various sources of systematic uncertainty affecting the measurement. Section 6.7 describes the kinematic fit used to determine the t tagging SFs and background normalizations. Finally, Section 6.8 describes how the measured data, predicted backgrounds, and fitted parame-

ters are used to construct a measured t jet p_T distribution, and the unfolding process used to extrapolate from the measured t jet p_T distribution to the cross section as a function of top quark p_T.

6.2 Data and Monte Carlo Samples

6.2.1 Data Sample

This measurement is performed at 13 TeV using the full 2016 CMS pp data sample. This data sample corresponds to an integrated luminosity of 35.9 fb^{-1}. Data are selected from the SingleMuon and SingleElectron HLT streams, using single muon and single electron triggers without lepton isolation criteria applied. The HLT path HLT_Ele45_CaloIdVT_GsfTrkIdT_PFJet200_PFJet50 is used for the e+jets channel, and the path HLT_Mu40_eta2p1_PFJet200_PFJet50 is used for the μ+jets channel. These triggers require an electron(muon) with $p_T > 45(40)$ GeV, a PF jet with $p_T > 200$ GeV, and a second PF jet with $p_T > 50$ GeV.

6.2.2 Simulated Samples

MC simulation is used to model the $t\bar{t}$ signal and most backgrounds. Signal $t\bar{t}$ events are simulated using POWHEG (v.2) [13–16] in combination with Pythia8 (v.8.205) [17–19]. POWHEG is used to model the hard process, using matrix elements calculated to NLO in QCD. Pythia8 simulates the parton shower and hadronization, using tune CUETP8M2T4 [20]. The simulation assumes a top quark mass of 172.5 GeV. The POWHEG simulation utilizes the NNPDF3.0 [76] PDF set. A priori, signal distributions are normalized to the NNLO inclusive $t\bar{t}$ cross section of 831.76 pb [77].

The dominant backgrounds to $t\bar{t}$ production in the semileptonic final state

include single top quark, W+jets, Z+jets, diboson, and QCD multijet production. Single top quark production contains contributions from t-channel and s-channel processes, as well as tW-channel processes where the top quark is produced in association with a W boson. Diboson backgrounds include contributions from WW, WZ, and ZZ production. QCD multijet background comprises events where jets are produced uniquely through strong interactions. POWHEG is used to simulate tW-channel and t-channel single top quark production, as well as diboson production. The MADGRAPH generator is used to simulate W+jets, Z+jets and QCD multijet processes. Additional partons in the final state are simulated at LO. The aMC@NLO generator is used to simulate s-channel single top production. The aMC@NLO and MADGRAPH generators are available through the MG5_AMC@NLO (2.2.2) package [22]. For all processes, PYTHIA8 is used to simulate the parton shower and hadronization, using the CUETP8M1 tune [23]. The MLM algorithm [24] is used to match the LO additional partons simulated by MADGRAPH with jets simulated by PYTHIA8. Separate W+jets and QCD multijet samples are generated for given ranges of H_T, the scalar sum of jet p_T. This improves the statistics of these samples in the high energy tail. For all samples, the detector response is simulated using GEANT4 [78]. A full list of the MC samples used in this analysis is provided in Table 6.1. A priori, simulated backgrounds are normalized to their theoretically predicted cross sections. The cross sections for the various background processes are also given in Table 6.1.

Corrections and Pileup Reweighting

Correction factors are applied to the simulated samples to match the conditions observed in data. The correction factors used to account for differences between simulated and actual trigger efficiencies, lepton ID and reconstruction efficiencies, and jet energy measurements are described in Chapter 5. In addition, simulated samples are corrected to accurately model the pileup conditions in data. Simulated samples are typically produced before data is collected, to allow quick comparisons between data and simulation. As such, simulated samples are produced with an estimated level of pileup, rather than the true level of pileup present in the data. In a given event, the expected number of pileup interactions is $N = \mathcal{L}\sigma$, where $\mathcal{L}$ is the instantaneous luminosity and $\sigma = 69.2$ mb is the total inelastic pp cross section. Integrating over the 2016 data set produces a distribution of the number of pileup interactions in data events. This distribution can be compared to the pileup distribution used in simulation. Dividing the data pileup distribution by the MC pileup distribution yields a pileup dependent SF, which can be used to reweight MC events to match pileup conditions in data.

Process	Dataset	N_{events}	σ [pb]
$t\bar{t}$	TT_TuneCUETP8M2T4_13TeV-powheg-pythia8	77229341	831.76
Single Top	ST_tW_top_5f_NoFullyHadronicDecays_13TeV-powheg_TuneCUETP8M1	8629641	19.3
	ST_tW_antitop_5f_NoFullyHadronicDecays_13TeV-powheg_TuneCUETP8M1	8681541	19.3
	ST_t-channel_top_4f_inclusiveDecays_13TeV-powhegV2-madspin-pythia8_TuneCUETP8M1	67240808	136
	ST_t-channel_antitop_4f_inclusiveDecays_13TeV-powhegV2-madspin-pythia8_TuneCUETP8M1	38811017	81
	ST_s-channel_4f_leptonDecays_13TeV-amcatnlo-pythia8	9651642	3.4
W+jets	WJetsToLNu_HT-100To200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	39617787	1345 * 1.21
	WJetsToLNu_HT-200To400_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	19914590	359.7 * 1.21
	WJetsToLNu_HT-400To600_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	5796237	48.91 * 1.21
	WJetsToLNu_HT-600To800_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	14822888	12.05 * 1.21
	WJetsToLNu_HT-800To1200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	6200954	5.501 * 1.21
	WJetsToLNu_HT-1200To2500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	6324934	1.329 * 1.21
	WJetsToLNu_HT-2500ToInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	2384260	0.0322 * 1.21
Z+jets	DYJetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	42923575	5765
Diboson	WW_TuneCUETP8M1_13TeV-pythia8	6987124	118.7
	WZ_TuneCUETP8M1_13TeV-pythia8	2995828	46.7
	ZZ_TuneCUETP8M1_13TeV-pythia8	990064	15.4
QCD	QCD_HT300to500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	17035891	347700
	QCD_HT500to700_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	18929951	32100
	QCD_HT700to1000_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	15629253	6831
	QCD_HT1000to1500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	4767100	1207
	QCD_HT1500to2000_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	3970819	119.9
	QCD_HT2000toInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	1991645	25.24

Table 6.1: MC samples used for the analysis, indicating the number of generated events and cross section for the generated process. Single top cross sections are calculated to NLO [79, 80], while Z+jets cross sections are calculated to NNLO [81]. W+jets and QCD cross sections as a function of H_T are calculated to LO [82], with the W+jets cross sections scaled by 1.21 to match the inclusive NLO value. The WW and ZZ cross sections are calculated to NNLO [83, 84], while the WZ cross section is calculated to NLO [85].

6.3 Event Selection

Boosted $t\bar{t}$ events in the semileptonic final state are expected to contain a t jet, b jet, muon or electron, and p_T^{miss} from the non-interacting neutrino. The $t\bar{t}$ pair is generally produced back-to-back, such that the b jet, lepton, and p_T^{miss} are in the opposite hemisphere from the t jet. This event topology provides a fairly clean $t\bar{t}$ signal, requiring little further selection to reduce backgrounds. As such, the event selection used to identify boosted $t\bar{t}$ events in the μ+jets and e+jets final states is mostly object based.

All events in the μ+jets channel must first pass a loose preselection, designed to be sensitive both to $t\bar{t}$ signal and most backgrounds. Events passing the preselection must contain exactly one signal muon and no signal electrons. At least one leptonic jet is required, where a leptonic jet is defined as an AK4 jet in the same hemisphere as the muon. Specifically, the leptonic jet must satisfy $0.3 < \Delta R(\mu, \text{jet}) < \pi/2$. At least one hadronic jet is required, where the hadronic jet is defined as an AK8 jet in the opposite hemisphere from the muon. Specifically, the hadronic jet must satisfy $\Delta R(\mu, \text{jet}) > \pi/2$. Finally, the event must have $p_T^{miss} > 35$ GeV. The signal lepton, AK4 jet, and AK8 jet definitions are provided in Chapter 5.

All events in the e+jets channel must also pass a loose preselection designed to be sensitive both to $t\bar{t}$ signal and most backgrounds. Events passing the preselection must contain exactly one signal electron and no signal muons. At least one leptonic jet is required, where a leptonic jet is defined as an AK4 jet satisfying $0.3 < \Delta R(e, \text{jet}) < \pi/2$. At least one hadronic jet is required, where the hadronic jet is defined as an AK8 jet satisfying $\Delta R(e, \text{jet}) > \pi/2$. The event must

have $p_T^{miss} > 50$ GeV. The p_T^{miss} is also required to satisfy

$$|\Delta\phi\left(X, p_T^{miss}\right) - 1.5| < 1.5 * \frac{p_T^{miss}}{110\ \text{GeV}} \tag{6.1}$$

where X is either the electron or the highest p_T jet. This eliminates events with high p_T^{miss} due to mismeasurement of the electron or jet p_T. The stricter requirements on p_T^{miss} in the electron channel are applied to improve rejection of QCD multijet background.

Events which pass the preselection are further categorized by whether the b or t jet candidate passes or fails the respective tagging criteria. The b jet candidate is the highest p_T leptonic jet in the event, while the t jet candidate is the highest p_T hadronic jet in the event. The tagging criteria are described in Chapter 5. Events are divided into the following exclusive kinematic regions:

- 0 t tag (0t): t jet candidate fails the t tagging requirement
- 1 t tag, 0 b tag (1t0b): t jet candidate passes the t tagging requirement, but b jet candidate fails the b tagging requirement
- 1 t tag, 1 b tag (1t1b): both t jet candidate and b jet candidate pass their respective tagging requirements

The kinematic regions are designed to have different admixtures of signal and background, with the 0t region being the most background dominated and the 1t1b region being the most signal dominated. These kinematic regions are used in the simultaneous likelihood fit to extract the background normalizations and t tagging SF.

6.4 Background Estimation

The main backgrounds to boosted semileptonic $t\bar{t}$ production are non-signal $t\bar{t}$, single top quark, W+jets, Z+jets, diboson, and QCD multijet production. Non-signal $t\bar{t}$ background includes dileptonic ($t\bar{t} \to WbWb \to \ell\nu b\ell\nu b$) and hadronic ($t\bar{t} \to WbWb \to q\bar{q}bq\bar{q}b$) $t\bar{t}$ decays, as well as τ+jets ($t\bar{t} \to WbWb \to \tau\nu bq\bar{q}b$) events where the τ decays to hadrons. Single top quark backgrounds includes tW-, t-, and s-channel production of a single top quark. In the tW channel, the top quark is produced in association with a W boson. W+jets and Z+jets backgrounds comprise production of a W or Z boson in association with additional partons. W+jets and Z+jets backgrounds mainly contribute to the semileptonic final state when the W or Z boson decays leptonically, via $W \to \ell\nu$ or $Z \to \ell\ell$. Diboson backgrounds include contributions from WW, WZ, and ZZ production. QCD multijet background events contain only jets produced purely through strong interactions. QCD multijet events least resemble the boosted semileptonic $t\bar{t}$ topology, but contribute to the final state due to the high QCD cross section. In general, none of the backgrounds truly produce a final state containing a t jet, b jet, lepton, and neutrino. Instead, the backgrounds contribute to the final state through jet or lepton misreconstruction. As such, the backgrounds can be mostly eliminated from the signal region using a sufficiently tight final state selection.

Non-signal $t\bar{t}$, single top quark, W+jets, Z+jets, and diboson backgrounds are modeled using MC. The same MC sample is used to model both non-signal and signal $t\bar{t}$, using knowledge of the true top quark decay to differentiate events. Separate W+jets samples are generated for different ranges of the scalar jet p_T sum H_T in order to improve simulation statistics at high energies. In contrast

to the other backgrounds, QCD multijet background is modeled using a data-driven technique described in Section 6.4.1. A QCD MC sample is used to cross-check the data control region used to extract the QCD shape. As with W+jets MC, the QCD MC sample is generated binned in H_T to improve statistics at high energies.

All MC samples used in the analysis are described in Table 6.1.

6.4.1 Data-driven QCD

The QCD multijet background is modeled by looking at data in a QCD dominated control region. The control region is orthogonal to the phase space covered by the preselection, which is referred to in this section as the signal region. The control region is defined by inverting the mini isolation requirement on the lepton in the preselection, requiring events to contain exactly one lepton with $0.1 < \text{miniIso} < 0.2$. The electron identification criteria are relaxed in the control region to improve statistics.

The number of observed data events in the control region, as well as the expected numbers of signal and background events from MC, are given in Table 6.2. The contribution from non-QCD processes in the control region is 28% in the μ+jets channel and 32% in the e+jets channel. The remaining events serve as the data-driven QCD estimate. This estimate agrees reasonably well with the QCD MC prediction.

The data-driven QCD is modeled by subtracting the expected contributions from $t\bar{t}$, single top, W+jets, Z+jets, and diboson processes from the observed

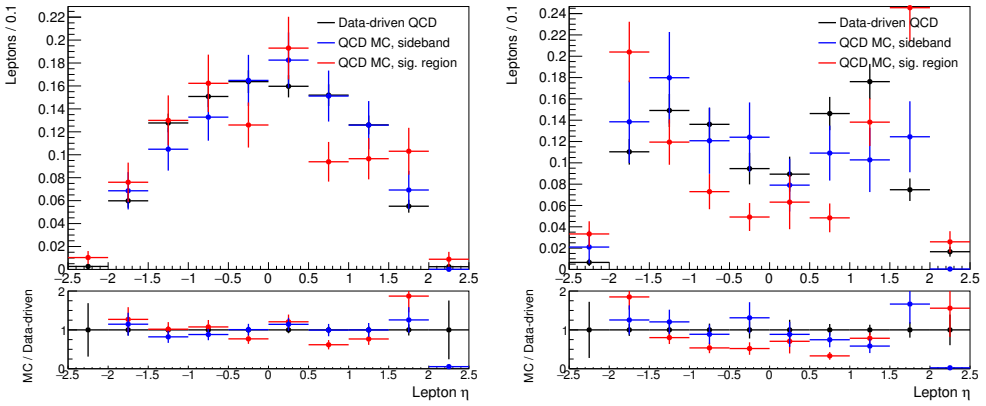
Sample	Event counts, μ+jets	Event counts, e+jets
$t\bar{t}$ (signal)	425 ± 13	176 ± 8
$t\bar{t}$ (non-semilep)	211 ± 9	87 ± 6
Single top	51 ± 2	20 ± 1
W+jets	320 ± 4	143 ± 3
Z+jets	30 ± 12	10 ± 7
Diboson	5 ± 2	4 ± 1
QCD (MC)	3397 ± 196	1255 ± 118
QCD (Data-driven)	2694	929
Data	3736	1369
% non-QCD in control region	28%	32%

Table 6.2: Expected signal and background yields, together with the observed number of events in data, in the QCD-dominated control region defined by inverting the isolation requirement on the lepton in the preselection. Uncertainties are statistical only.

data in the control region. To verify that the QCD kinematics resulting from the data-driven method appropriately describe the QCD kinematics in the signal region, the data-driven QCD kinematics are compared to the QCD MC expectation in both the signal and control regions. Comparing data-driven QCD to QCD MC in the control region indicates whether the QCD MC provides an accurate description of the data. Comparing the QCD distributions in the control region to those in the signal region demonstrates whether there are any systematic shifts in the QCD kinematics between the two regions. If all three shapes are reasonably consistent, the data-driven QCD is expected to accurately model the QCD kinematics in the signal region.

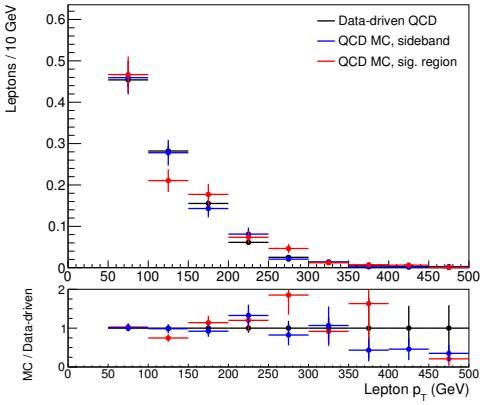
Selected comparisons of the data-driven QCD, control region QCD MC, and signal region QCD MC kinematics are shown in Figure 6.1. In general, there is reasonable agreement between the MC prediction for QCD in the control region and the data-driven QCD estimate in the control region. The QCD MC is therefore assumed to provide a roughly accurate description of the QCD kine-

matics. However, some differences are seen between the QCD kinematics in the signal and control regions. A systematic shift is observed in the electron η and p_T distributions, where electrons in the signal region are more likely to have high p_T and $|\eta|$ than electrons in the control region. The source of the disagreement may be traced to the mini isolation definition. As described in Appendix A, the electron mini isolation cone is hollow in the endcap but not in the barrel. In QCD multijet events, objects reconstructed as electrons are most likely misidentified jets. At high p_T, a light jet misidentified as an electron could become narrow enough to fall within the inner radius of the isolation cone in the endcap. The light jet would then appear to be fully isolated. Thus, high p_T fake electrons in the endcap will contribute to the signal region rather than the control region. Despite this issue, the data-driven technique is judged to give the best description of QCD kinematics. The QCD contribution is therefore modeled using kinematic shapes from the data-driven technique, normalized to a zeroth-order estimate of the QCD normalization taken from QCD MC.

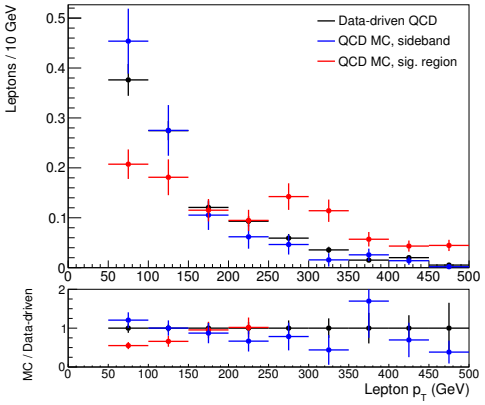


(a) μ+jets

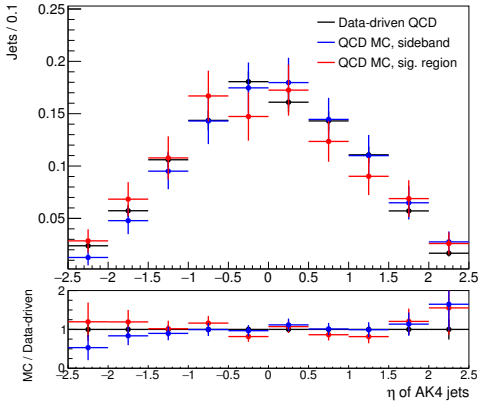
(b) e+jets



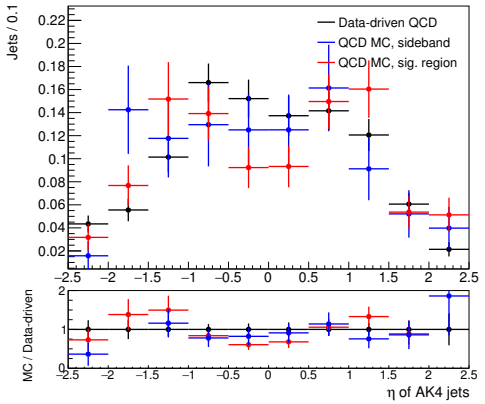
(c) μ+jets



(d) e+jets



(e) μ+jets



(f) e+jets

Figure 6.1: Comparisons of QCD kinematics as predicted by MC in the signal region, MC in the control region (sideband), and the data-driven technique. Lepton η distributions are shown on top, lepton p_T in the middle, and b jet candidate η on the bottom; the μ+jets channel is shown on the left and e+jets on the right. The largest difference is between the electron η and p_T distributions in the signal and control regions.

6.5 Comparison of Data and MC Prediction

This section presents a comparison between the observed data and the expected contributions from signal and background. The comparison includes event yields in the 0t, 1t0b, and 1t1b kinematic regions, as well as selected kinematic distributions in each region. The backgrounds, and the methods used to model them, are described in more detail in Section 6.4. The background normalizations, lepton SFs, and t tagging SFs used to produce these comparisons are not the final values used in the $t\bar{t}$ cross section measurement. The final values of these parameters are determined through a kinematic fit to data, as described in Section 6.7. The comparisons shown here, using a priori parameter values, serve as a check of the modeling of the kinematic variables which are used for the fit.

6.5.1 Event Yields

The a priori event yields in each kinematic region are listed in Table 6.3 for the μ+jets channel and in Table 6.4 for the e+jets channel. The quoted uncertainties are statistical only. The event yields are determined by scaling the number of simulated events which pass the selection to the expected number of events in data, using the integrated luminosity and cross section for each process. Specifically, the expected number of events in data for a given process is

$$\mathrm{N_{exp}} = \mathrm{N_{sel}} * \frac{\mathcal{L} * \sigma}{\mathrm{N_{tot}}}, \tag{6.2}$$

where N_{gen} is the total number of simulated events for the process in ques-

tion, N_{sel} is the number of simulated events for that process which pass the selection, $\mathcal{L}$ is the integrated luminosity, and σ is the process cross section.

Sample	Number of events (μ+jets channel)			
	Preselection	0t	1t0b	1t1b
$t\bar{t}$ (signal)	26580 ± 102	17344 ± 118	4837 ± 73	4399 ± 41
$t\bar{t}$ (non-signal)	4748 ± 43	4146 ± 46	317 ± 19	284 ± 10
Single top	3257 ± 16	2871 ± 17	250 ± 6	136 ± 3
W+jets	30860 ± 43	27157 ± 46	3565 ± 15	138 ± 3
Z+jets	3048 ± 121	2714 ± 128	309 ± 42	26 ± 11
Diboson	596 ± 19	553 ± 19	40 ± 5	3 ± 1
QCD	2917 ± 70	2681 ± 67	182 ± 19	54 ± 8
Total	72006 ± 185	57466 ± 199	9500 ± 89	5040 ± 45
Data	60672	49137	7348	4187

Table 6.3: Estimates of the signal and background yields, together with the observed number of events in data, in each kinematic region for the μ+jets channel. The quoted uncertainties are statistical only.

Sample	Number of events (e+jets channel)			
	Preselection	0t	1t0b	1t1b
$t\bar{t}$ (signal)	16627 ± 80	10508 ± 93	3191 ± 59	2928 ± 33
$t\bar{t}$ (non-signal)	3479 ± 37	3042 ± 39	246 ± 15	190 ± 8
Single top	2201 ± 13	1943 ± 14	165 ± 5	93 ± 3
W+jets	18293 ± 35	16075 ± 37	2137 ± 12	81 ± 2
Z+jets	1225 ± 76	1047 ± 82	163 ± 31	16 ± 9
Diboson	390 ± 15	361 ± 16	27 ± 4	2 ± 1
QCD	2605 ± 110	2339 ± 102	253 ± 41	13 ± 4
Total	44819 ± 165	35314 ± 171	6182 ± 81	3323 ± 36
Data	39313	31559	4801	2953

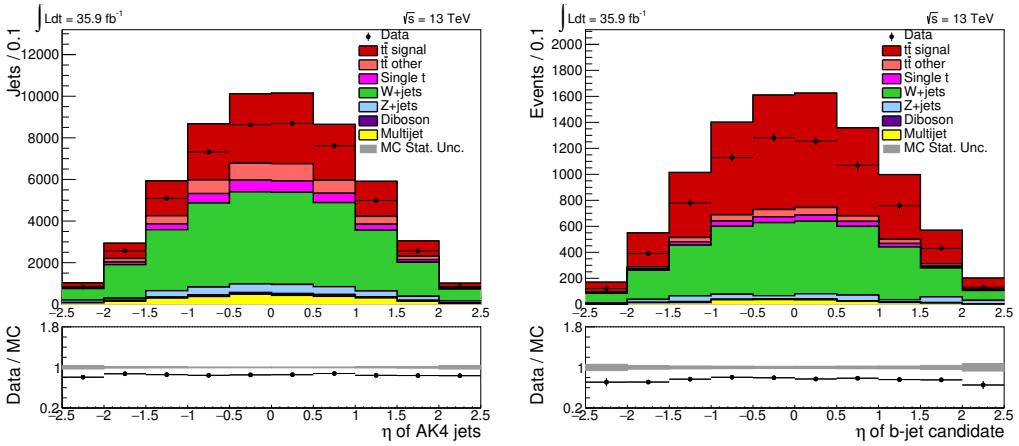
Table 6.4: Estimates of the signal and background yields, together with the observed number of events in data, in each kinematic region for the e+jets channel. The quoted uncertainties are statistical only.

6.5.2 Kinematic Distributions

A variety of kinematic distributions are studied to compare data to the signal and background predictions. Selected distributions, namely those used in the kinematic fit and those used to extract the differential cross section, are shown in this section. Kinematic distributions are shown separately for events in the 0t, 1t0b, and 1t1b regions. All figures shown in this section use the a priori normalizations given in Tables 6.3 and 6.4.

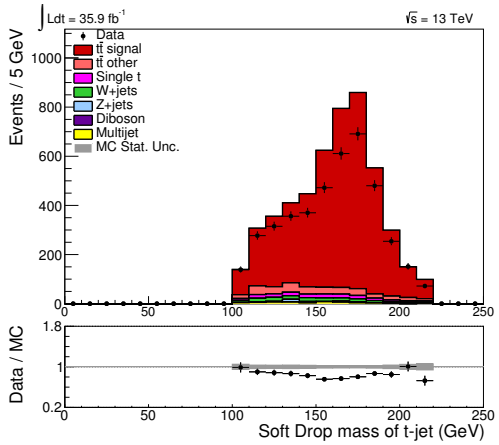
The kinematic distributions used in the likelihood fit are the b jet candidate η in the 0t and 1t0b regions and the invariant mass of the soft drop groomed t-tagged jet in the 1t1b region. The a priori versions of these distributions are shown in Figures 6.2 and 6.3 for the μ+jets and e+jets channels, respectively. The Monte Carlo predictions are seen to overestimate the data in terms of the yield. Figures 6.4 and 6.5 show the p_T spectrum for the t jet candidate in each of the three regions for the μ+jets and e+jets channels.

The kinematic distributions demonstrate that the 0t region is dominated by W+jets background, with significant contributions from QCD multijet and $t\bar{t}$ signal. The 1t0b region is dominated by $t\bar{t}$ signal, but still contains a significant amount of W+jets background. The 1t1b region is fully dominated by the $t\bar{t}$ signal, with only small background contributions from QCD, non-signal $t\bar{t}$, W+jets, and single top.



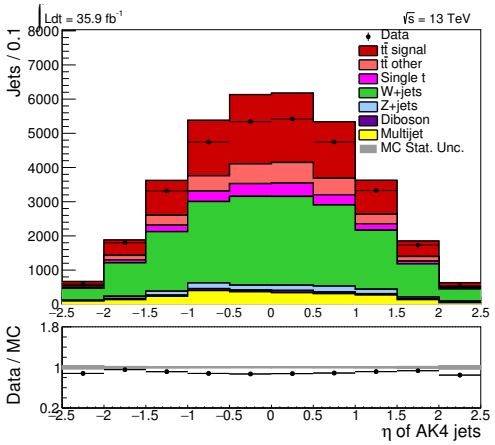
(a) 0t

(b) 1t0b

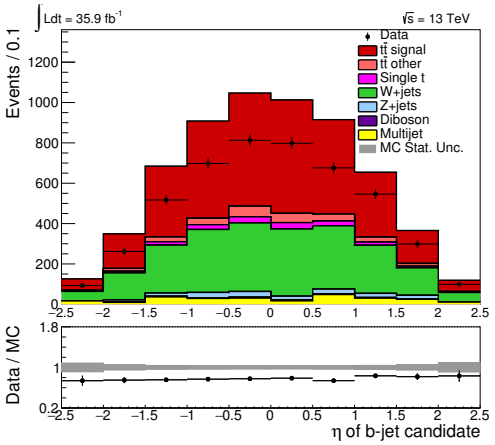


(c) 1t1b

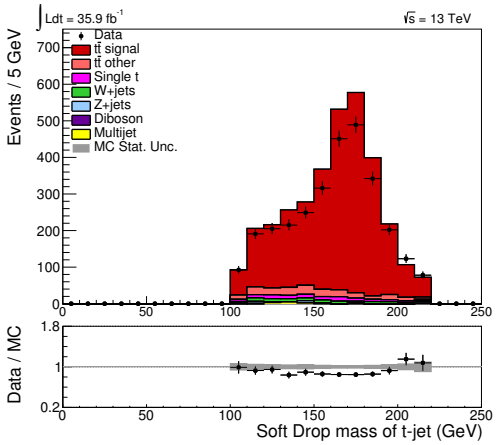
Figure 6.2: A priori kinematic distributions used in the likelihood fit, for events in the μ+jets channel. The b jet candidate η is used for the 0t and 1t0b regions, while the invariant mass of the t jet after soft drop grooming is used in the 1t1b region.



(a) 0t

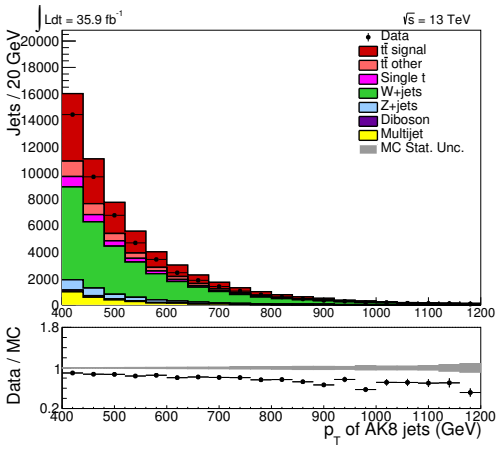


(b) 1t0b

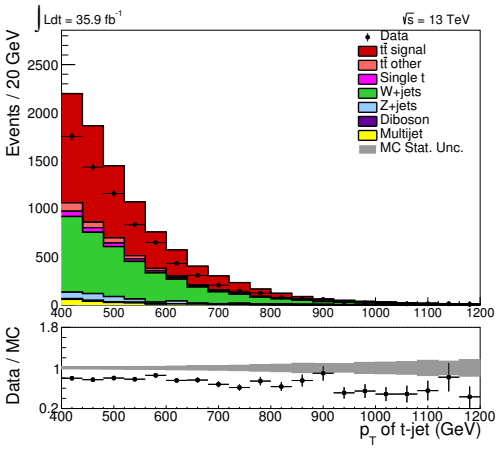


(c) 1t1b

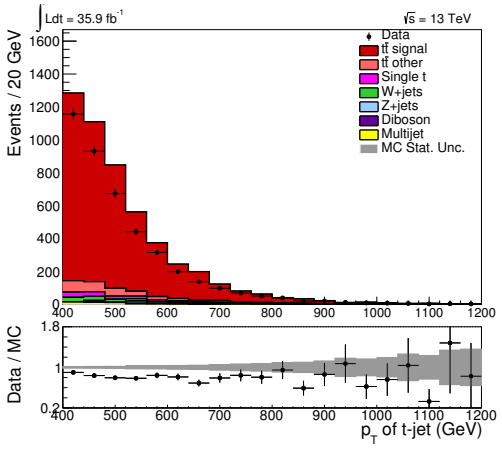
Figure 6.3: A priori kinematic distributions used in the likelihood fit, for events in the e+jets channel. The b jet candidate η is used for the 0t and 1t0b regions, while the invariant mass of the t jet after soft drop grooming is used in the 1t1b region.



(a) 0t

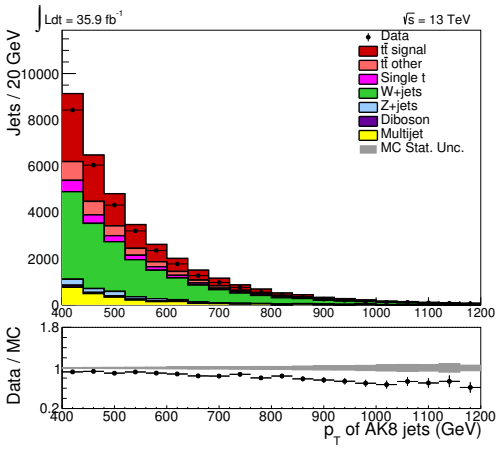


(b) 1t0b

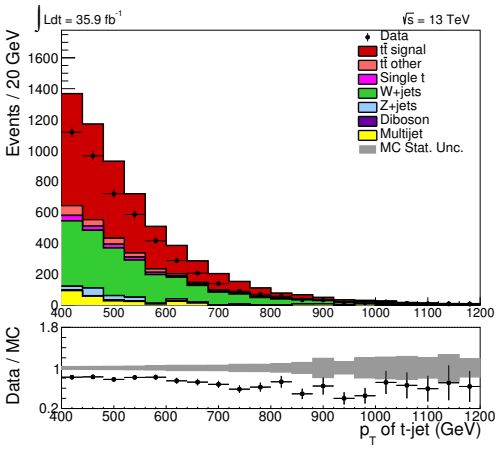


(c) 1t1b

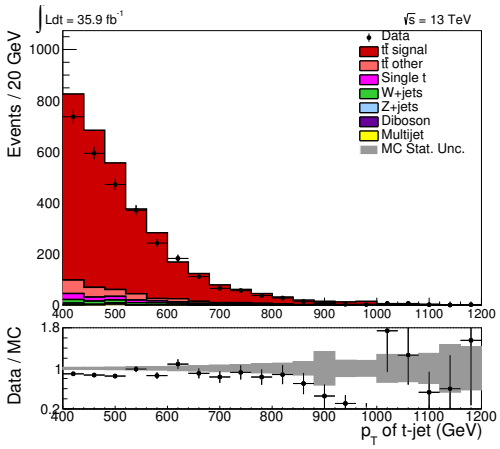
Figure 6.4: A priori p_T of the t jet candidate in the 0t, 1t0b, and 1t1b regions, for events in the μ+jets channel



(a) 0t



(b) 1t0b



(c) 1t1b

Figure 6.5: A priori p_T of the t jet candidate in the 0t, 1t0b, and 1t1b regions, for events in the e+jets channel

6.6 Systematic Uncertainties

Many sources of systematic uncertainty can impact the $t\bar{t}$ cross section measurement. Experimental systematic uncertainties affect the identification and reconstruction of $t\bar{t}$ events, while theoretical uncertainties affect the extraction of the $t\bar{t}$ cross section from observed events. Systematic uncertainties may affect the number of $t\bar{t}$ events or the $t\bar{t}$ event kinematics. Most systematic uncertainties affecting CMS measurements are well known and modeled using dedicated measurements. However, some systematic uncertainties affecting the boosted semileptonic $t\bar{t}$ cross section measurement are unknown or poorly modeled. Such uncertainties are initially assigned conservative estimates, which may be constrained in the kinematic fit to data described in Section 6.7.3.

6.6.1 Experimental Uncertainties

A number of experimental uncertainties affect the $t\bar{t}$ cross section measurement. Among these, the most dominant are the uncertainties on the b tagging and t tagging SFs, which describe differences in the tagging efficiencies between simulation and data. The various sources of experimental uncertainty are described below, in roughly decreasing order of impact.

When measuring the $t\bar{t}$ cross section in the boosted semileptonic final state, the $t\bar{t}$ cross section and t tagging efficiency are highly correlated and difficult to measure independently. The efficiency of the t tagging algorithm used in this analysis has been independently measured in a phase space very similar to the signal region of this measurement [86]. However, the efficiency measurement

implicitly assumes that the $t\bar{t}$ MC provides an accurate simulation of $t\bar{t}$ production in the measurement phase space. Using this efficiency would therefore introduce a bias in our measurement of the $t\bar{t}$ cross section. Instead, the t tagging SF is extracted simultaneously with the integrated $t\bar{t}$ cross section using a kinematic fit. The fit procedure is described in detail in Section 6.7. To parameterize our initial ignorance of the t tagging efficiency SF, the a priori t tagging SF is set to 1.0 with an uncertainty of 25%. Separate, uncorrelated t tagging SFs describe the tagging of true t jets and the mistagging of fake t jets. The t tag SF is applied to the $t\bar{t}$ and single top quark samples, while the t mistag SF is applied to the W+jets, Z+jets, diboson, and QCD multijet samples. The impact of the t tagging SF uncertainty is assessed by producing kinematic distributions with the t tagging SF varied within uncertainties.

The b tagging SFs are measured as a function of the b jet p_T and η [61] using tag-and-probe techniques. Tag-and-probe techniques exploit two-component final states, where the identification of one object (the tag) implicitly identifies the second component (the probe). The true identity of the probe object is thus assumed to be known, despite no selection being directly applied. This is comparable to the way that the true identity of a MC object can be known by looking at the simulation record. Selection efficiencies measured using probe objects are thus considered the true selection efficiencies in data, suitable for comparison to true selection efficiencies in MC. Tagging SFs are evaluated for true b jets, while mistagging SFs are evaluated for u, d, s, and c jets. MC events are weighted according to whether the b jet candidate passes or fails the b tagging requirement, as described in 5.6. The impact of the b tagging SF uncertainty is assessed by producing kinematic distributions with the b tagging SFs varied within uncertainties. The b tag and b mistag SFs for all flavours are varied up or down

simultaneously, effectively treating the tag and mistag SFs as fully correlated.

The uncertainty on the difference between the true and reconstructed energies of jets is parameterized as an uncertainty on the jet energy corrections (JEC) described in Section 5.4. These uncertainties are measured along with the JEC in 2016 data [59]. The JEC uncertainties depend on the corrected jet p_T and uncorrected jet direction. The impact of the JEC uncertainty σ is assessed by scaling each jet four-momentum by $(1 \pm \sigma)$ and producing new kinematic distributions using the scaled jets. While the JEC uncertainties are different for the AK4 and AK8 jets, the impact of the uncertainties is evaluated in the same fashion. The AK4 and AK8 JEC are varied up and down together, treating the two uncertainties as fully correlated.

The jet energy resolution (JER) is known to be worse in data than predicted by simulation. To account for this, the momenta of simulated jets are smeared to match the JER in data [58]. Uncertainties on the JER affect the width of this smearing function. The impact of uncertainties on the JER is determined by producing kinematic distributions where the width of the JER smearing function has been varied within uncertainties.

To account for differences between the simulated amount of pileup and the pileup observed in data, MC events are reweighted to reproduce the expected data pileup distribution. The expected data pileup distribution is calculated assuming a total pp inelastic cross section of 69.2 mb. The uncertainty on the total pp inelastic cross section is ±4.6 %. To assess the impact of this uncertainty on the expected data pileup distribution, and hence the pileup reweighting applied to MC, pileup distributions are calculated where the total pp inelastic cross section is varied within its uncertainty. These shifted pileup distributions are then

used to reweight the MC.

The performance of the lepton identification, reconstruction, trigger and isolation provides a small source of uncertainty. The SFs for the muon identification, reconstruction, and trigger efficiencies are measured using tag-and-probe techniques [54,87]. The total muon SF uncertainty is assessed by simultaneously varying the various muon SFs up or down within their respective uncertainties, treating the different SF uncertainties as fully correlated. The SFs for the electron track reconstruction, identification, isolation, and trigger efficiencies are also measured using tag-and-probe techniques [51, 87]. The total electron SF uncertainty is assessed by simultaneously varying the various electron SFs up or down within their respective uncertainties, treating the different SF uncertainties as fully correlated. The lepton ID, trigger, and reconstruction efficiency SFs are jointly referred to as the lepton SFs.

A conservative estimate of the uncertainty on the background rates yields an ad-hoc initial uncertainty of 30% on the single top quark, W+jets, Z+jets, and diboson rates. An additional uncertainty on the flavor composition of the jets in the W+jets background is assessed by dividing the simulated W+jets samples into W+heavy flavour and W+light flavor components using MC truth information. In the W+heavy flavor sample, W bosons are produced in association with one or two b or c quarks. The normalizations of the W+light flavor and W+heavy flavor components are allowed to vary separately with 30% uncertainty. A 50% uncertainty is imposed on the QCD rate, to reflect the combined uncertainty on the QCD production rate and the extraction of the kinematic distributions from the sideband.

As described in Section 6.5.1, the integrated luminosity is used to determine

the expected number of events to pass a selection in data, given the number of events which pass the selection in simulation. Uncertainties on the integrated luminosity thus affect the total number of expected events in data. The uncertainty on the integrated luminosity is measured to be 2.5% [88].

6.6.2 Theoretical Uncertainties

A number of theoretical uncertainties affect the MC modeling of $t\bar{t}$ events, and hence the SM measurements extracted from observation of $t\bar{t}$ processes. These uncertainties generally affect the parameters used by the simulation to describe the hard process, parton shower, and hadronization. Some theoretical uncertainties are assessed through generating alternative MC samples with shifted simulation parameters, while other theoretical uncertainties are assessed through reweighting the nominal simulated sample. The main theoretical uncertainties affecting this measurement are described below. The treatment of these theoretical uncertainties generally follows the prescriptions detailed in [89].

The energy scales μ_R and μ_F related to renormalization and factorization affect the simulation of the semileptonic $t\bar{t}$ decay by POWHEG. Uncertainties on μ_R and μ_F are assessed by calculating event weights corresponding to variations of μ_R and μ_F. Specifically, weights are calculated for independent variations of μ_R and μ_F by [2.0,0.5]. The unphysical anticorrelated variations are discarded, yielding a total of seven weights corresponding to seven combinations of the renormalization and factorization scales. The total scale uncertainty is defined as the envelope of the seven $[\mu_R, \mu_F]$ variations, according to the prescription

in [90]. The effect of the scale uncertainty is assessed by varying the event weight within the scale uncertainty.

Uncertainties in the energy scale also affect the strength of the strong coupling α_S, which affects the generation of ISR and FSR by POWHEG. The impact of uncertainties in α_S^{ISR} and α_S^{FSR} is assessed by generating samples with α_S^{ISR} and α_S^{FSR} varied within their uncertainties. Samples for the α_S^{ISR} up, α_S^{ISR} down, α_S^{FSR} up, and α_S^{FSR} down variations are generated separately. After these samples were produced, data from LEP was found to constrain the FSR uncertainty by a factor of $\sqrt{2}$. Because of this, differences between the nominal sample and the samples with α_S^{FSR} varied up or down are scaled down by $\sqrt{2}$.

The parton distribution function (PDF) used by the generator describes the identity and energies of the partons participating in the hard collision. The POWHEG sample is simulated using the NNPDF3.0 PDF set [76]. To model the effects of uncertainties in the PDF, 100 MC replicas are drawn from the PDF probability distribution. These replicas cover the envelope of the PDF uncertainties. For each replica, event weights are calculated to parameterize the probability of generating the same event with the shifted PDF. The standard deviation of these weights is used as the PDF uncertainty. The effect of the PDF uncertainty is assessed by varying the event weight within the PDF uncertainty.

The separate samples used to describe theory uncertainties are listed in Table 6.5. The systematic variation samples have lower statistics than the nominal $t\bar{t}$ MC sample.

Systematic	Parameter	Variation	Sample	N_{events}
Shower Scales	α_S^{ISR}	Up	TT_TuneCUETP8M2T4_13TeV-powheg-isrup-pythia8	59033604
		Down	TT_TuneCUETP8M2T4_13TeV-powheg-isrdown-pythia8	58999580
	α_S^{FSR}	Up	TT_TuneCUETP8M2T4_13TeV-powheg-fsrup-pythia8	59230899
		Down	TT_TuneCUETP8M2T4_13TeV-powheg-fsrdown-pythia8	59306906

Table 6.5: MC samples used for theory systematic variations

6.7 Kinematic Fit

The first step in measuring the differential $t\bar{t}$ cross section as a function of top quark p_T is to measure the t jet p_T distribution for $t\bar{t}$ events. The measured t jet p_T distribution is determined through defining a signal region in data dominated by $t\bar{t}$ events, and removing the expected contribution from background sources. The subtracted backgrounds are weighted with appropriate scale factors, including a t tagging SF. This section describes the kinematic fit used to simultaneously determine the t tag SF, lepton SFs, and background normalizations used to construct the measured t jet p_T distribution.

6.7.1 Fit Mechanism

The fit is performed as a simultaneous binned likelihood fit in the exclusive 0t, 1t0b, and 1t1b kinematic regions. The three regions are designed such that the t tagging efficiency is constrained by the relative population of events in each region. The varying admixtures of signal and background between the different regions also offer more power to determine the background rates. The fit is implemented as a likelihood maximization, using a statistical tool developed for the combination of CMS and ATLAS Higgs search results [91]. Given an expectation $Y(\alpha)$ as a function of parameters α and an observation X, the likelihood of $Y(\alpha)$ being the correct description of X is the same expression as the probability of observing X given expectation $Y(\alpha)$. The maximum likelihood estimator $\hat{\alpha}$ is the set of parameters which maximize the likelihood, corresponding to the most likely parameters given the observed data. In actual practice, $\hat{\alpha}$ is usually determined by minimizing the negative logarithm of the likelihood (NLL). For

a simultaneous binned likelihood fit, the NLL takes the form

$$\text{NLL} = -log\ \text{L} = -\sum_{c}\sum_{i} log\left[\frac{y_{ic}(\alpha)^{x_{ic}}e^{y_{ic}(\alpha)}}{x_{ic}!}\right] + \text{constraint terms} \tag{6.3}$$

where c is the channel or kinematic region, i is the bin number, x_{ic} is the observed number of events in bin i of channel c, and y_{ic} is the expected number of events in bin i of channel c. The expected number of events in a given bin and channel receives contributions from signal and non-signal $t\bar{t}$, single top quark, W+jets, Z+jets, diboson, and QCD multijet backgrounds. The resulting expression for $y_{ic}(\alpha)$ is

$$y_{ic}(\alpha) = N_{sc}(\alpha)s_{ic}(\alpha) + \sum_{b} N_{bc}(\alpha)b_{ic}(\alpha) \tag{6.4}$$

where N_{sc} is the number of signal events expected in channel c, s_{ic} is the relative fraction of signal events expected in bin i of channel c, N_{bc} is the number of events expected from background b in channel c, and b_{ic} is the relative fraction of background b events expected in bin i of channel c. The parameter of interest in the fit is the signal strength r, which modifies the signal rate in the fit. The remaining parameters in α are a set of nuisance parameters ρ corresponding to systematic uncertainties. Each systematic uncertainty is constrained, and contributes a constraint term to the NLL.

6.7.2 Treatment of Systematic Uncertainties

For the purposes of the fit, this analysis contains two types of systematic uncertainties. The first type, rate uncertainties, includes uncertainties on the single top quark, W+jets, Z+jets, diboson, and QCD production rates. These systematics only affect the overall normalization of each background, not the shape of the kinematics used in the fit. Since the rate uncertainties are multiplicative, they are expressed as lognormal distributions scaling the a priori background rate. The expected number of background events, as a function of the background rate uncertainty nuisance parameter ρ_b, is given as

$$N_{bc}(\rho_b) = N_{bc0}(1 + \Delta_b)^{\rho_b} \tag{6.5}$$

where N_{bc0} is the MC expectation and Δ_b is the a priori relative uncertainty on the rate. Each nuisance parameter ρ_b is distributed according to a gaussian with mean 0 and standard deviation 1, contributing a constraint term of $e^{\rho_b^2}$ to the NLL.

The second set of uncertainties, shape uncertainties, affect both the total number of expected events and the kinematic shape. These uncertainties include the lepton SF, jet energy scale, jet energy resolution, b tag SF, and t tag SF uncertainties. To describe the effect of these systematics, kinematic distributions are provided which correspond to $\pm 1\sigma$ variations of the relevant systematic. An interpolation procedure is used to determine the kinematic distributions for other values of the systematic. Separate interpolations are used to determine the normalization and normalized shape of the kinematic distribution. Since the dependence of the kinematic shape on a given parameter is likely both com-

plicated and unknown, a quadratic interpolation is used for systematic shifts below 1σ, and linear extrapolation used for shifts beyond 1σ. The template normalizations are interpolated asymmetrically and linearly in log scale, with positive systematic shifts determined through interpolating the nominal and $+1\sigma$ normalizations and negative systematic shifts determined through interpolating the nominal and -1σ normalizations. The nuisance parameters themselves continue to be normally distributed, contributing constraint terms of $e^{\rho_s^2}$

Systematic uncertainties can be specified to affect specific channels and samples in the fit. Systematic uncertainties which behave differently for different samples, channels, or regions are assigned separate nuisance parameters for these regions. In this case, the effects of the systematic uncertainty in these regions are fully uncorrelated. The single top quark, W+heavy flavor, W+light flavor, Z+jets, and diboson rate uncertainties apply only to their respective samples and are applied in all regions and channels. For the purpose of the fit W+heavy flavor and W+light flavor events are treated as separate background sources, although the two backgrounds may be combined into a single W+jets sample when presenting event counts or kinematic distributions. Separate nuisance parameters describe the QCD rate uncertainty in the μ+jets and e+jets channels, since the QCD kinematics are expected to be different in each channel. The jet energy scale, jet energy resolution, and b tag SF nuisance parameters apply to all samples, regions, and channels. In regions where the b jet candidate fails the tagging requirement, the same b tag SF nuisance parameter describes the b mistag SF. Separate nuisance parameters are used to describe the t tag and t mistag SF uncertainties. The t tag SF nuisance parameter is applied to the $t\bar{t}$ and single top quark samples, where a true t jet is expected, while the t mistag SF nuisance parameter is applied to the remaining samples. Separate nuisance

parameters describe the lepton SF uncertainties in the μ+jets and e+jets channels. Finally, the single parameter of interest is the signal strength r used to scale the signal $t\bar{t}$ rate.

The uncertainties described above do not include the theoretical uncertainties described in Section 6.6.2. These theoretical uncertainties are evaluated independently from the kinematic fit.

6.7.3 Fit Results

The kinematic fit is performed using different kinematic variables in each region to discriminate $t\bar{t}$ signal from background. The b jet candidate η is used in the 0t and 1t0b regions while the soft drop groomed invariant mass of the t-tagged jet is used in the 1t1b region. These distribtuions were chosen to offer good discrimination between $t\bar{t}$, W+jets, and QCD processes. Figures 6.2 and 6.3 show the kinematic distributions prior to the fit in the μ+jets and e+jets channels. The signal and non-signal $t\bar{t}$ distributions are merged into a single distribution for the purpose of the fit, essentially constraining the $t\bar{t}$ semileptonic branching ratio to be exactly equal to that provided by MC simulation.

The kinematic fit is first performed separately for the individual μ+jets and e+jets channels. To combine the channels, the kinematic fit is also evaluated by fitting μ+jets and e+jets events together, with all nuisance parameters generally constrained to be the same in both channels. As mentioned previously, separate nuisance parameters are used for the muon and electron SFs, as well as the muon and electron QCD rates.

The posterior kinematic distributions in the three fit categories are shown for the μ+jets channel fit in Figure 6.6 and for the e+jets channel fit in Figure 6.7. The posterior kinematic distributions for the combined fit in both the μ+jets and e+jets channels are shown in Figures 6.8 and 6.9, respectively.

In addition to determining the most likely values of the signal strength and the nuisance parameters, the kinematic fit measures the correlations between parameters. The fitted values of the nuisance parameters and their correlations are shown for the μ+jets channel fit in Figure 6.10, for the e+jets channel fit in Figure 6.11, and for the combined fit in Figure 6.12. Since the nuisance parameters are initially normally distributed, with mean 0 and variance 1, a posterior nuisance parameter with central value 0 and uncertainty 1 is not constrained by the fit.

In general, most of the posterior nuisance parameters show agreement with their a priori values. The nuisance parameters most constrained by the fit are the t tag and t mistag SFs, as expected. A slight tension between the μ+jets and e+jets channels leads to greater disagreement between the prior and posterior nuisance parameters in the combined fit. The largest (anti)correlations in the single-channel fits are between the signal strength and t tag SF, t tag SF and t mistag SF, W+light flavor and W+heavy flavor rates, and signal strength and single top quark rate. In the combined fit there is also a correlation between the muon and electron SFs, likely due to the slight tension between the μ+jets and e+jets channels.

The posterior event counts which account for all posterior parameters are given in Table 6.6 for the fit in the μ+jets channel. Table 6.7 gives the posterior event counts for the e+jets channel, and Table 6.8 the posterior event counts in

both channels for the combined fit.

Figures 6.13 and 6.14 show the p_T spectrum for the t jet candidate in each of the three regions for the μ+jets and e+jets channels, respectively. These distributions use the posterior t tag SF, lepton SF, and background normalizations, but not the posterior JER or JEC, or b tag SF.

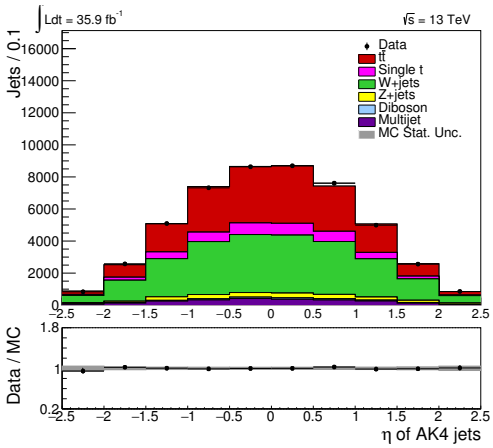
Multiple cross-checks were performed to assess the fit stability, as described in detail in [87]. As a result of these checks, additional systematic uncertainties were assigned to the posterior t tag SF, lepton SF, and background normalizations.

Sample	Number of events (μ+jets channel)		
	0t	1t0b	1t1b
$t\bar{t}$	18275 ± 1392	4236 ± 166	3841 ± 96
Single top quark	3958 ± 1092	330 ± 106	182 ± 57
W+jets	22151 ± 2864	2449 ± 336	105 ± 20
Z+jets	1895 ± 371	180 ± 46	15 ± 7
Diboson	548 ± 165	33 ± 11	2 ± 1
QCD multijet	2315 ± 891	127 ± 56	38 ± 18
Total	49141 ± 3506	7356 ± 396	4182 ± 115
Data	49137	7348	4187

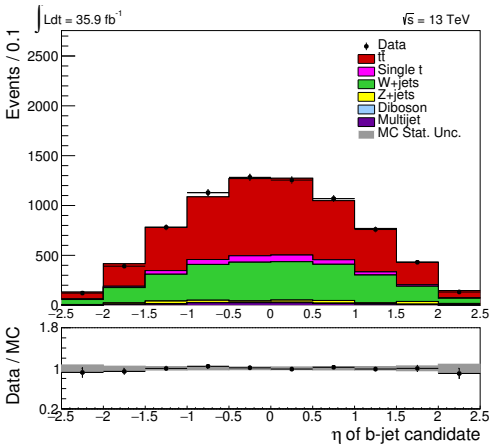
Table 6.6: Post-fit signal and background event yields in the 0t, 1t0b, and 1t1b regions, together with observation in data, for the fit in the μ+jets channel. The $t\bar{t}$ yields include both signal and non-signal contributions, while the W+jets yields include both W+light flavor and W+heavy flavor contributions. The uncertainties include all post-fit experimental uncertainties.

6.7.4 Interpretation of Results

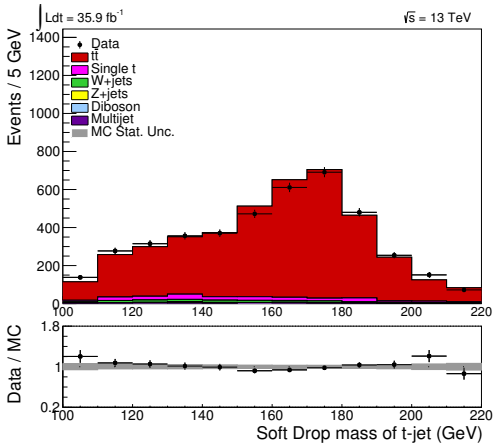
The main goal of the kinematic fit is to simultaneously determine the parameters needed to construct the measured t jet p_T distribution from background-



(a) η of b jet candidate, 0t



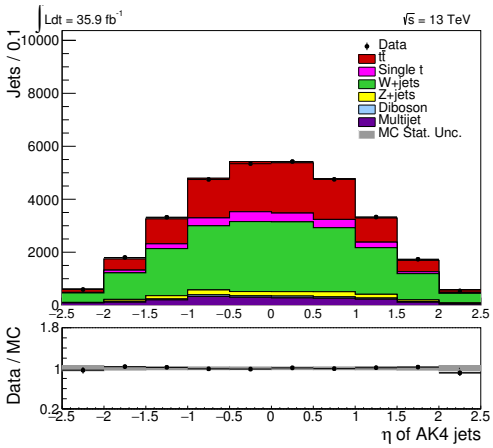
(b) η of b jet candidate, 1t0b



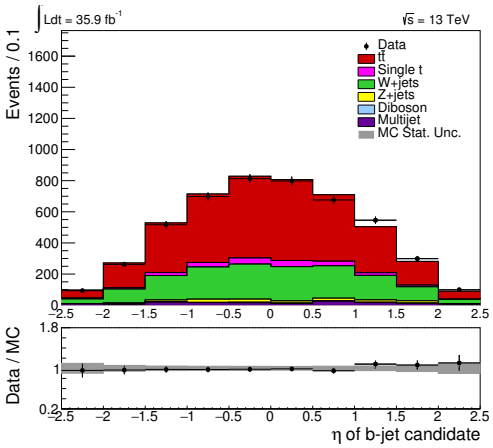
(c) groomed t jet invariant mass, 1t1b

Figure 6.6: Post-fit kinematic distributions for the fit in the μ+jets channel. The posterior agreement between data and MC is within uncertainties.

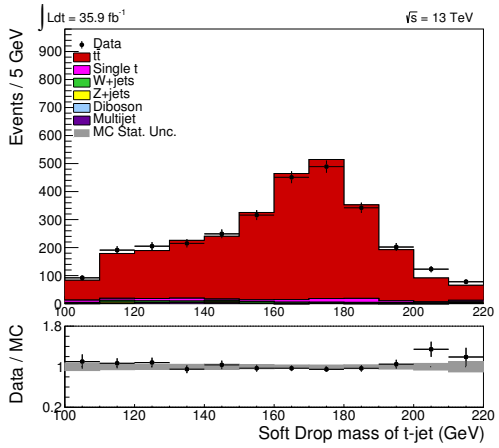
subtracted data. Specifically, the posterior t tag SF, lepton SFs, and background normalizations are used to construct the background-subtracted data distribution. The a priori values of the JER, JEC, and b tag SFs are used as these values have been independently measured with high precision. While the fit effectively extracts the integrated $t\bar{t}$ cross section for top quarks with $p_T > 400$ GeV, by means of the posterior signal strength r applied to the a priori $t\bar{t}$ cross section, this value is not directly used in the later cross section measurement. However,



(a) η of b jet candidate, 0t



(b) η of b jet candidate, 1t0b



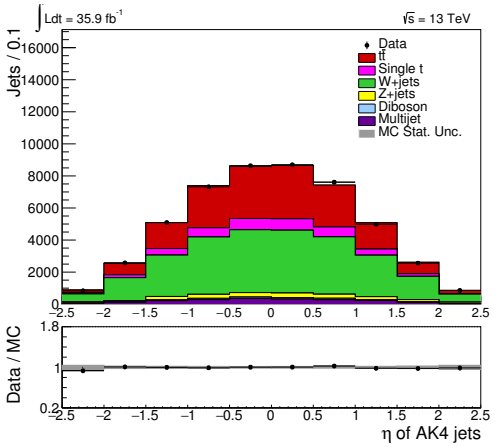
(c) groomed t jet invariant mass, 1t1b

Figure 6.7: Post-fit kinematic distributions for the fit in the e+jets channel. The posterior agreement between data and MC is within uncertainties.

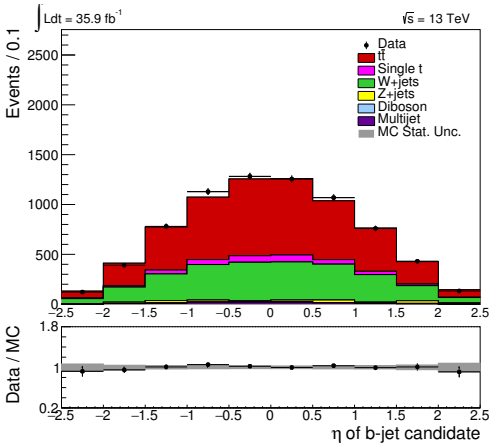
the signal strength r does provide a useful estimator of the overall fit result.

Posterior t Tagging Scale Factor

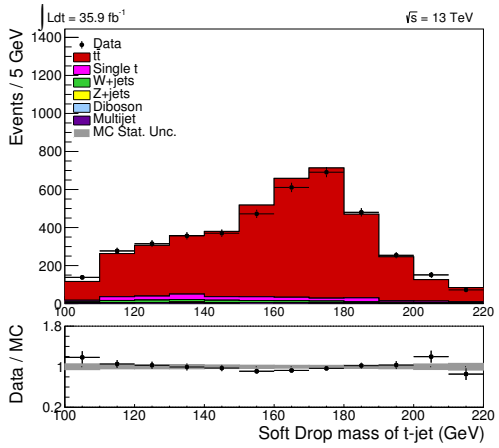
The posterior nuisance parameters for the t tag and t mistag scale factors are given in Table 6.9 for the μ+jets channel, e+jets channel, and combined fits. There is some tension between the results in the e+jets and μ+jets channels, but



(a) η of b jet candidate, 0t



(b) η of b jet candidate, 1t0b

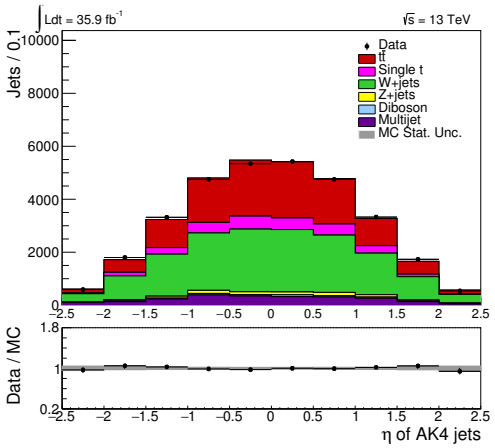


(c) groomed t jet invariant mass, 1t1b

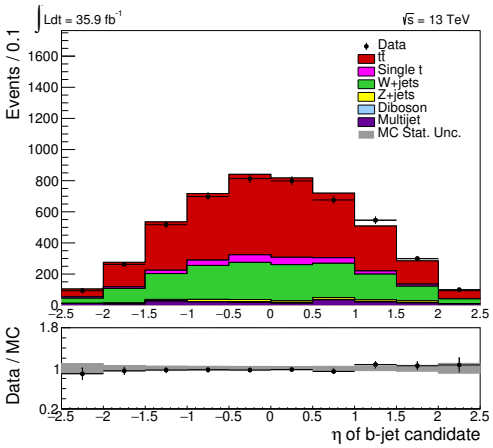
Figure 6.8: Post-fit kinematic distributions for the combined fit, for events in the μ+jets channel. The posterior agreement between data and MC is within uncertainties.

the results bracket and agree with the combined result. The posterior nuisance parameters from the combined fit are used to determine the posterior t tag SF and t mistag SF.

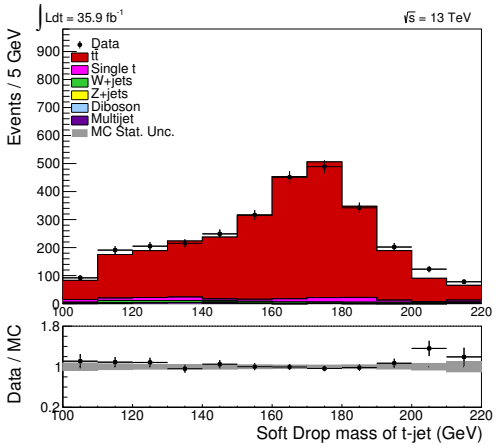
In addition to the uncertainty from the fit, an additional systematic uncertainty on the t tag and t mistag SF nuisance parameters is assessed as a result of a set of fit cross checks, described in detail in [87]. The final posterior nuisance



(a) η of b jet candidate, 0t



(b) η of b jet candidate, 1t0b

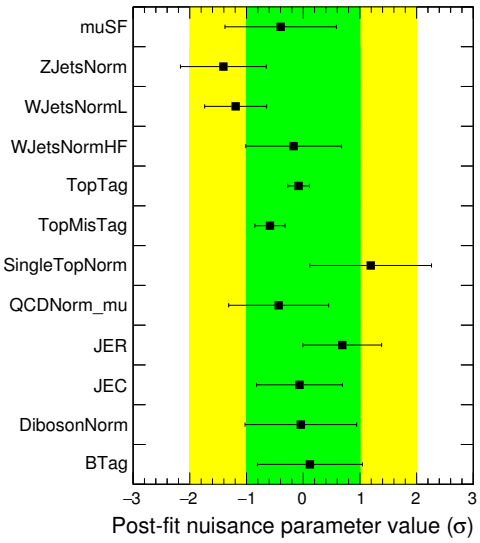


(c) groomed t jet invariant mass, 1t1b

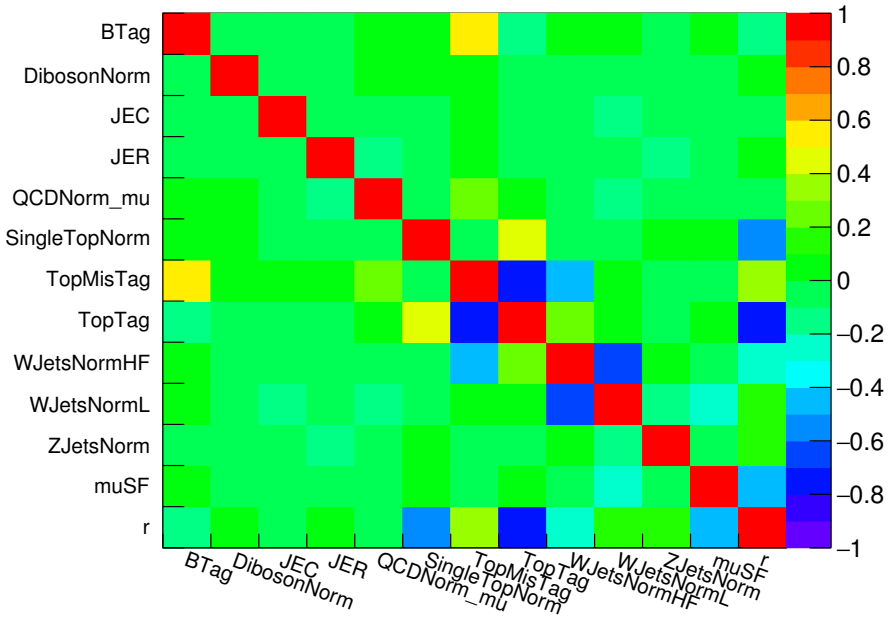
Figure 6.9: Post-fit kinematic distributions for the combined fit, for events in the e+jets channel. The posterior agreement between data and MC is within uncertainties.

parameters, along with the corresponding posterior t tag and t mistag SFs, are given in Table 6.10.

The t tag SF is independently calculated to be $1.06^{+0.09}_{-0.04}$ in a phase space very similar to our signal region [86]. The t tag SF we calculate shows very good agreement, and comparable uncertainty, with this value.

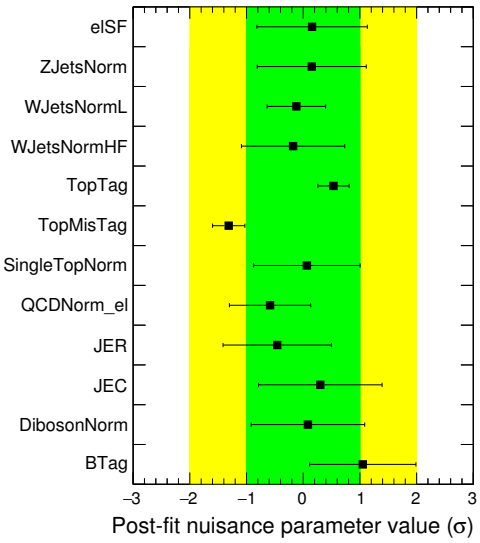


(a)

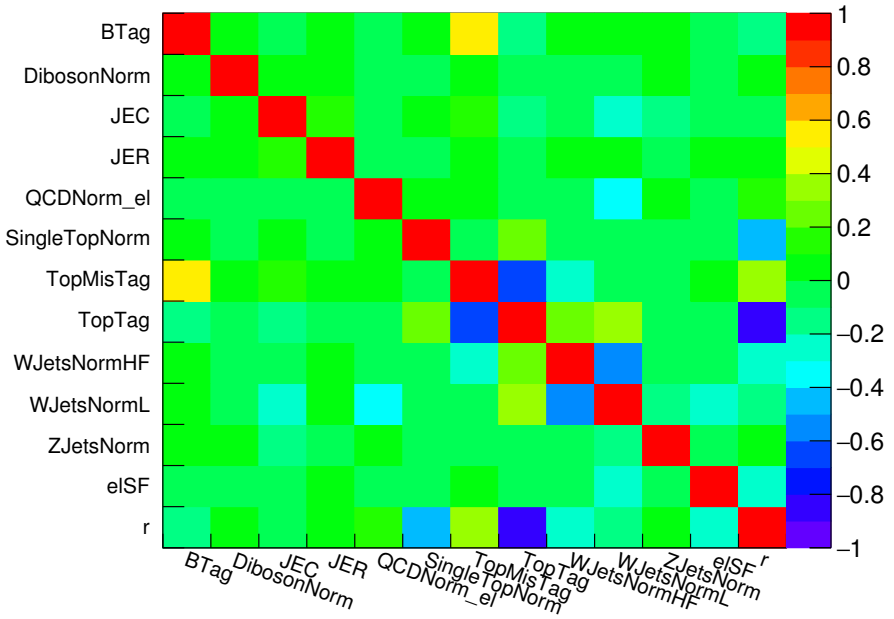


(b)

Figure 6.10: Post-fit nuisance parameters (a) and their correlations (b) for the fit in the μ+jets channel. Nuisance parameters compatible with zero indicate no modification to the systematic by the fit; nuisance parameter uncertainties of 1σ indicate no constraint on the systematic uncertainty due to the fit.

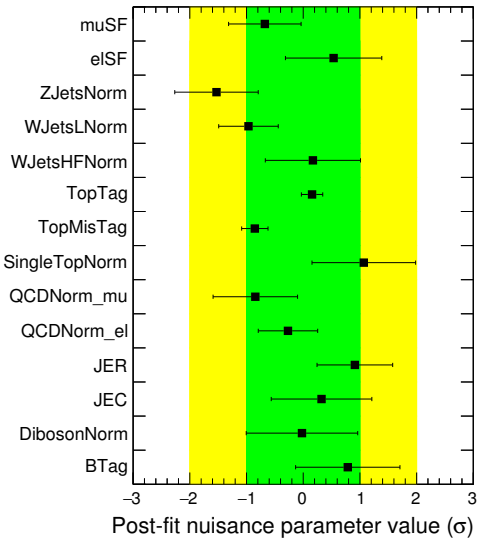


(a)

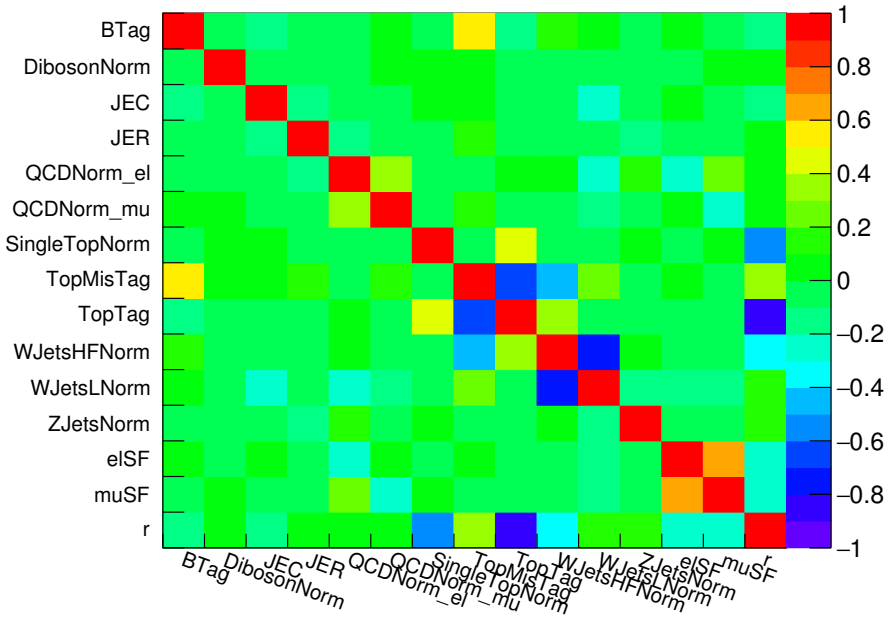


(b)

Figure 6.11: Post-fit nuisance parameters (a) and their correlations (b) for the fit in the e+jets channel. Nuisance parameters compatible with zero indicate no modification to the systematic by the fit; nuisance parameter uncertainties of 1σ indicate no constraint on the systematic uncertainty due to the fit.

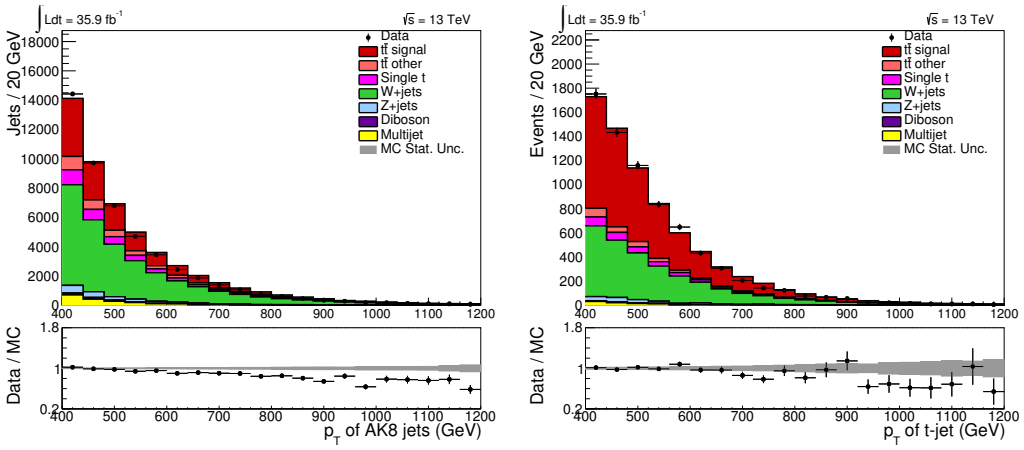


(a)



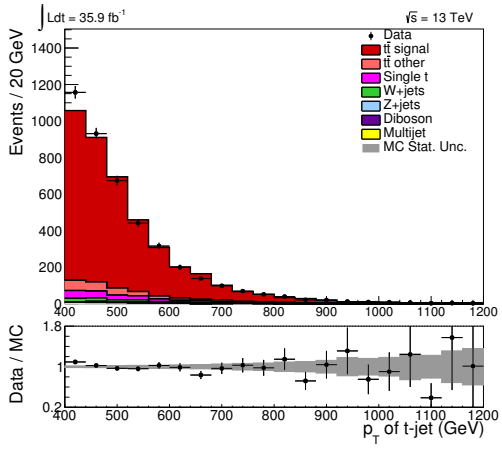
(b)

Figure 6.12: Post-fit nuisance parameters (a) and their correlations (b) for the fit in the combined channel. Nuisance parameters compatible with zero indicate no modification to the systematic by the fit; nuisance parameter uncertainties of 1σ indicate no constraint on the systematic uncertainty due to the fit.



(a) 0t

(b) 1t0b

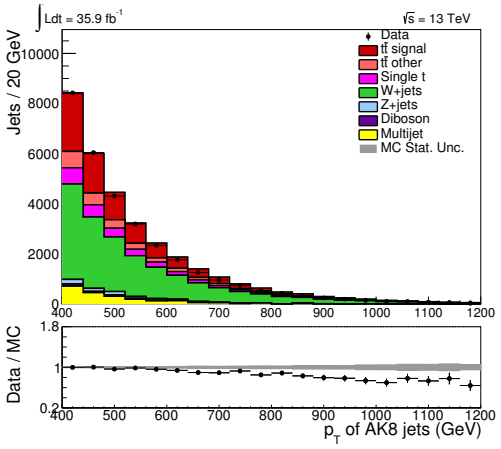


(c) 1t1b

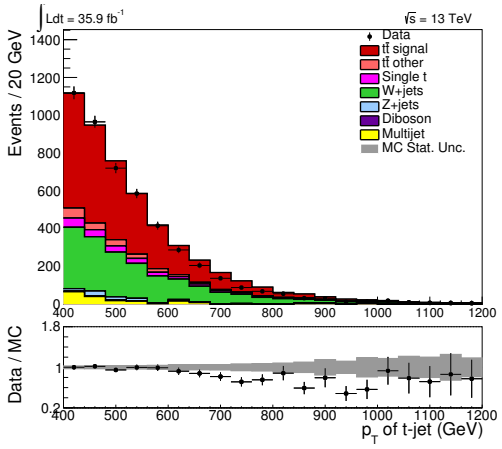
Figure 6.13: p_T of the t jet candidate in the 0t, 1t0b, and 1t1b regions, for events in the μ+jets channel. Posterior t tag SF, muon SF, and background normalizations are used.

Posterior lepton scale factors

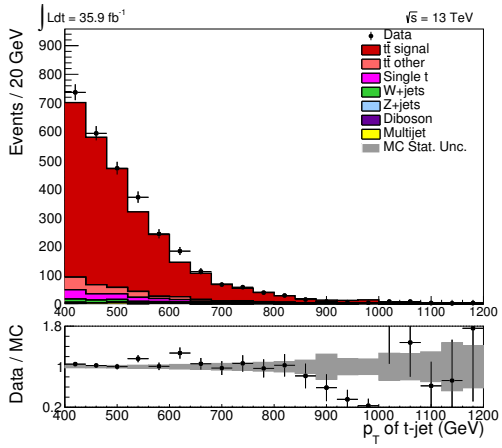
The posterior nuisance parameters for the muon and electron combined ID, reconstruction, and isolation SFs are given in Table 6.11 for the μ+jets channel, e+jets channel, and combined fits. The posterior nuisance parameters from the combined fit are used to determine the posterior muon and electron SFs.



(a) 0t



(b) 1t0b



(c) 1t1b

Figure 6.14: p_T of the t jet candidate in the 0t, 1t0b, and 1t1b regions, for events in the e+jets channel. Posterior t tag SF, electron SF, and background normalizations are used.

In addition to the uncertainty from the fit, an additional systematic uncertainty on the muon and electron SF nuisance parameters is assessed as a result of a set of fit cross checks, described in detail in [87]. The final posterior nuisance parameters are given in Table 6.12. These nuisance parameters are used to determine the posterior lepton SF, using the interpolation of the SFs corresponding to nuisance parameters of 0, 1, and -1.

Sample	Number of events (e+jets channel)		
	0t	1t0b	1t1b
$t\bar{t}$	9783 ± 959	2943 ± 114	2748 ± 70
Single top quark	1926 ± 485	199 ± 57	97 ± 29
W+jets	16371 ± 2056	1417 ± 195	57 ± 11
Z+jets	1160 ± 299	117 ± 39	12 ± 8
Diboson	382 ± 119	19 ± 7	2 ± 1
QCD multijet	1937 ± 583	136 ± 51	8 ± 6
Total	31559 ± 2414	4831 ± 242	2924 ± 77
Data	31559	4801	2953

Table 6.7: Post-fit signal and background event yields in the 0t, 1t0b, and 1t1b regions, together with observation in data, for the fit in the e+jets channel. The $t\bar{t}$ yields include both signal and nonsignal contributions, while the W+jets yields include both W+light flavor and W+heavy flavor contributions. The uncertainties include all post-fit experimental uncertainties.

Posterior background normalizations

The posterior nuisance parameters for the background rates are given in Table 6.13 for the μ+jets channel, e+jets channel, and combined fits. While the nuisance parameters determined from the separate fits do not necessarily bracket the nuisance parameters determined in the combined fit, the results all agree within uncertainties. The posterior nuisance parameters from the combined fit are used to determine the posterior background rates.

In addition to the uncertainty from the fit, an additional systematic uncertainty on the background normalization nuisance parameters is assessed as a result of the fit checks performed in [87]. The final posterior nuisance parameters, along with the resulting ratio of posterior to prior background rate, are given in Table 6.14.

Sample	Number of events (μ+jets channel)		
	0t	1t0b	1t1b
$t\bar{t}$	16911 ± 1227	4217 ± 170	3896 ± 88
Single top quark	3825 ± 950	339 ± 100	189 ± 55
W+jets	24027 ± 2737	2432 ± 292	110 ± 19
Z+jets	1851 ± 373	162 ± 41	14 ± 7
Diboson	552 ± 143	30 ± 9	2 ± 1
QCD multijet	2001 ± 615	99 ± 34	30 ± 11
Total	49168 ± 3231	7279 ± 356	4241 ± 107
Data	49137	7348	4187

Sample	Number of events (e+jets channel)		
	0t	1t0b	1t1b
$t\bar{t}$	10945 ± 820	2878 ± 118	2672 ± 67
Single top quark	2575 ± 632	241 ± 70	117 ± 36
W+jets	14638 ± 1669	1495 ± 181	65 ± 11
Z+jets	738 ± 155	88 ± 25	9 ± 6
Diboson	373 ± 96	21 ± 7	2 ± 1
QCD multijet	2267 ± 522	172 ± 52	10 ± 7
Total	31534 ± 2040	4896 ± 234	2876 ± 77
Data	31559	4801	2953

Table 6.8: Post-fit signal and background event yields in the 0t, 1t0b, and 1t1b regions, together with observation in data, for the fit in the combined channel. The $t\bar{t}$ yields include both signal and nonsignal contributions, while the W+jets yields include both W+light flavor and W+heavy flavor contributions. The uncertainties include all post-fit experimental uncertainties.

Signal Strength

The posterior signal strength describes the relation between the measured inclusive $t\bar{t}$ cross section and the MC prediction. While the signal strength is not directly used to construct the measured t jet p_T distribution, it is implicitly related to the normalization of this distribution. Therefore, the signal strength remains a useful indicator of the overall fit result. The signal strengths are given in Table 6.15 for the μ+jets channel, e+jets channel, and combined fits. In addition to the uncertainty from the fit, additional systematic uncertainties on the

Fit	Posterior nuisance parameters	
	t tag SF	t mistag SF
μ+jets channel	-0.08 ± 0.19	-0.59 ± 0.27
e+jets channel	0.53 ± 0.28	-1.31 ± 0.29
Combined	0.16 ± 0.19	-0.85 ± 0.23

Table 6.9: Posterior t tag and t mistag SF nuisance parameters given by the fits in the μ+jets, e+jets, and combined channels. There is reasonable agreement between the results.

	Nuisance parameter	Fitted SF
t tag SF	0.16 ± 0.19 (fit) ± 0.27 (sys)	1.04 ± 0.08
t mistag SF	-0.85 ± 0.23 (fit) ± 0.21 (sys)	0.83 ± 0.06

Table 6.10: Posterior t tag and t mistag SF nuisance parameters with total uncertainty, along with corresponding posterior t tag and t mistag SFs. The a priori t tag and t mistag SFs had a central value of 1.0 and relative uncertainty of 25%.

Fit	Posterior nuisance parameters	
	Muon SF	Electron SF
μ+jets channel	-0.40 ± 0.98	N/A
e+jets channel	N/A	0.16 ± 0.97
Combined	-0.68 ± 0.64	0.54 ± 0.85

Table 6.11: Posterior muon and electron SF nuisance parameters given by the fits in the μ+jets, e+jets, and combined channels. There is reasonable agreement between the results.

	Nuisance parameter
Muon SF	-0.68 ± 0.64 (fit) ± 0.72 (sys)
Electron SF	0.54 ± 0.85 (fit) ± 0.28 (sys)

Table 6.12: Posterior muon and electron SF nuisance parameters with total uncertainty.

Sample	Posterior Nuisance Parameter		
	μ+jets channel fit	e+jets channel fit	Combined fit
Single top quark	1.19 ± 1.07	0.06 ± 0.94	1.07 ± 0.91
W+light flavor	-1.19 ± 0.55	-0.12 ± 0.52	-0.97 ± 0.53
W+heavy flavor	-0.17 ± 0.85	-0.18 ± 0.91	0.17 ± 0.84
Z+jets	-1.41 ± 0.76	0.15 ± 0.96	-1.53 ± 0.74
Diboson	-0.04 ± 0.99	0.08 ± 1.00	-0.02 ± 0.98
QCD (μ+jets)	-0.43 ± 0.88	N/A	-0.84 ± 0.75
QCD (e+jets)	N/A	-0.58 ± 0.72	-0.27 ± 0.52

Table 6.13: Posterior background normalization nuisance parameters given by the fits in the μ+jets, e+jets, and combined channels. There is reasonable agreement between the results in the different channels.

Sample	Nuisance parameter	Posterior factor
Single top quark	1.07 ± 0.91 (fit) ± 1.01 (sys)	$1.32^{+0.57}_{-0.39}$
W+light flavor	-0.97 ± 0.53 (fit) ± 0.63 (sys)	$0.78^{+0.18}_{-0.15}$
W+heavy flavor	0.17 ± 0.84 (fit) ± 0.48 (sys)	$1.05^{+0.30}_{-0.24}$
Z+jets	-1.53 ± 0.74 (fit) ± 0.90 (sys)	$0.67^{+0.24}_{-0.18}$
Diboson	-0.02 ± 0.98 (fit) ± 0.28 (sys)	$0.99^{+0.31}_{-0.23}$
QCD (μ+jets)	-0.84 ± 0.75 (fit) ± 1.61 (sys)	$0.71^{+0.75}_{-0.36}$
QCD (e+jets)	-0.27 ± 0.52 (fit) ± 1.25 (sys)	$0.90^{+0.65}_{-0.38}$

Table 6.14: Posterior background normalization nuisance parameters with total uncertainty, along with corresponding posterior normalization factors. The normalization factors, applied to a priori background normalizations, yield the posterior normalizations. The a priori background normalization uncertainties were 30% for the single top quark, W+light flavor, W+heavy flavor, Z+jets, and diboson components, and 50% for the QCD multijet components.

signal strength are assessed as a result of a series of fit cross checks, as described in [87]. The measured signal strengths indicate that MC overpredicts the rate of $t\bar{t}$ production in the boosted regime.

Fit	Signal strength
μ+jets channel	0.84 ± 0.06 (fit) ± 0.10 (sys)
e+jets channel	0.77 ± 0.05 (fit) ± 0.07 (sys)
Combined	0.80 ± 0.04 (fit) ± 0.07 (sys)

Table 6.15: Signal strength given by the fits in the μ+jets, e+jets, and combined channels. There is reasonable agreement between the results.

6.8 Measurement of Differential Cross Section

In this section, a measurement of the differential $t\bar{t}$ cross section is presented. Specifically, the cross section is measured for the inclusive process $pp \rightarrow t\bar{t}$ as a function of the top quark p_T. Since this cross section only describes the hard interaction which generates the $t\bar{t}$ pair, without imposing any requirements on the top quark decay or interaction with the detector, the phase space is referred to as the full parton level phase space.

In scenarios with an imperfect detector, the number of times a process is observed $\mathrm{N_{obs}}$ is related to the process cross section σ via

$$\mathrm{N_{obs}} = \sigma \mathcal{L} \mathrm{A} \epsilon \tag{6.6}$$

where $\mathcal{L}$ is the integrated luminosity, A the geometric detector acceptance, and ϵ the event reconstruction efficiency. The acceptance accounts for events which physically fall outside the detector coverage, while the efficiency accounts for events which are improperly reconstructed by the detector. In the case of a differential cross section measurement, the relation between the cross section and number of observed events is more complicated. In our case,

$$\mathrm{N_{obs}}\left(p_{\mathrm{T}}^{\mathrm{t\ jet}}\right) = \mathcal{L} \int \frac{d\sigma}{dp_{\mathrm{T}}^{\mathrm{t}}} * \epsilon\left(p_{\mathrm{T}}^{\mathrm{t}},\ p_{\mathrm{T}}^{\mathrm{t\ jet}}\right) * A\left(p_{\mathrm{T}}^{\mathrm{t}},\ p_{\mathrm{T}}^{\mathrm{t\ jet}}\right) * dp_{\mathrm{T}}^{\mathrm{t}} \tag{6.7}$$

where the dependence of the observation on the probability of reconstructing a t jet with $p_T^{\mathrm{t\ jet}}$ given a true top quark with p_T^{t} is made explicit. Determining $d\sigma/dp_{\mathrm{T}}$ as a function of the top quark p_T, when given an observed number of

events as a function of t jet p_T, requires this relation to be inverted through an unfolding procedure.

The observed number of events as a function of t jet p_T is extracted from the signal-dominated 1t1b region. Expected background contributions are subtracted from the observed data, using the background normalizations extracted from the kinematic fit as given in Table 6.8. The posterior t tagging SFs and lepton SFs determined by the combined fit are applied to the MC in place of their a priori values. This background-subtracted data distribution is considered the measured distribution of t jet p_T for $t\bar{t}$ events. This measured distribution is then unfolded to extract the $t\bar{t}$ cross section in bins of top quark p_T. The unfolding relies on a response matrix mapping the p_T of the t-tagged jet to the top quark p_T, essentially parameterizing $\epsilon\left(p_{\mathrm{T}}^{\mathrm{t}},\ p_{\mathrm{T}}^{\mathrm{t\ jet}}\right) * A\left(p_{\mathrm{T}}^{\mathrm{t}},\ p_{\mathrm{T}}^{\mathrm{t\ jet}}\right)$. This response matrix is derived from $t\bar{t}$ MC.

The systematic uncertainty on the unfolded measurement receives contributions from both experimental and theoretical sources. The experimental uncertainties, described in Section 6.6.1, include uncertainties on the lepton SFs, luminosity, pileup corrections, jet energy scale, jet energy resolution, b tagging SFs, t tagging SFs, and background rates. The posterior values from the kinematic fit are used for the t tagging SFs, background rates, and lepton SFs, while the a priori values are used for the remaining uncertainties. The theoretical uncertainties on the PDF, renormalization and factorization scales, and ISR / FSR shower scales are also included in the cross section measurement.

The unfolding is performed separately in the μ+jets and e+jets channels, as well as for the combined lepton+jets measurement.

6.8.1 Parton Level Phase Space

While the cross section is measured for the full parton level phase space, the unfolding procedure actually extrapolates the number of observed $t\bar{t}$ events in the lepton+jets final state to the semileptonic parton level phase space where the $t\bar{t}$ decay proceeds as $t\bar{t} \to WbWb \to \ell\nu bq\bar{q}b$, with $\ell = e, \mu$. In this way, the unfolding procedure only accounts for the detector acceptance and reconstruction efficiencies of true semileptonic events. The cross section is then extrapolated from the semileptonic parton level phase space to the full parton level phase space using independently calculated branching ratios for the semileptonic $t\bar{t}$ decay.

In order to correctly align the $t\bar{t}$ cross section measurement in the parton level phase space with theoretical predictions, the parton level objects must be defined coherently in simulation and theoretical calculations. In simulation, the evolution of the collision output is described by the event record. The event record contains a list of all particles produced in the event, where each particle has a unique identifier, an ID describing the particle type, a status code, a four-momentum, and information on the mother and daughter particles. Since the event record includes all intermediate states, a given particle may have multiple entries describing its evolving kinematics. The entry which corresponds to the particle state in the theoretical calculation is identified by looking at the status, mother, and daughter information of the simulated particle.

Semileptonic events are selected at parton level in the following way. First, the event is required to contain exactly one prompt lepton, meaning the lepton is produced in the hard interaction rather than during the parton shower. The lepton mother information is used to determine whether the lepton is prompt. Prompt leptons may be electrons, muons, or taus, where the particle flavor is

given by the particle ID. Prompt electrons and muons are required to be final state particles, as described by the particle status code. Prompt taus are required to decay to a final state muon or electron.

The parton level top quark is defined as the top quark post FSR and prior to decaying. Since the observation is performed as a function of the t jet p_T, the parton level top quark is chosen to be the top quark which will later decay hadronically. Assuming a semileptonic final state, the hadronically decaying top quark is the antitop quark if the final state lepton is positively charged and the top quark if the final state lepton is negatively charged. The parton level top quark p_T is required to be above 400 GeV, matching the cut on the reconstructed t jet p_T.

6.8.2 Improving Unfolding Statistics

Since the unfolding procedure is very sensitive to statistical fluctuations, special care is used to improve the statistics of the simulated samples used to perform the unfolding. In addition to the inclusive $t\bar{t}$ MC sample used for the kinematic fit, dedicated samples generated with high $t\bar{t}$ invariant mass are used to improve statistics in the high-energy tail. Samples are generated for 700 GeV $< M_{t\bar{t}} <$ 1000 GeV and 1000 GeV $< M_{t\bar{t}}$, where $M_{t\bar{t}}$ refers to the invariant mass of the top quark pair produced by the generator prior to FSR. In addition, an extension to the inclusive $t\bar{t}$ sample is used to improve statistics at lower energies. These samples are described in Table 6.16. To prevent overlap between the samples, a cut of $M_{t\bar{t}} <$ 700 GeV is applied to the inclusive $t\bar{t}$ samples.

To combine the three $M_{t\bar{t}}$ ranges, each sample is reweighted to match the

expected number of observed events in data. The samples each receive a weight

$$w = \frac{f\sigma\mathcal{L}}{\mathrm{N_{tot}}} \tag{6.8}$$

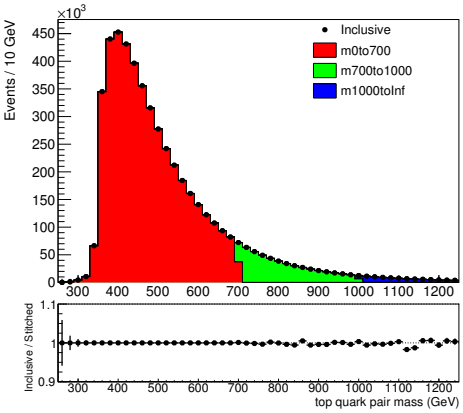
where σ is the inclusive $t\bar{t}$ cross section, f is the fraction of $t\bar{t}$ events in the given mass range, $\mathcal{L}$ is the integrated luminosity, and N_{tot} is the total number of simulated events in the sample. The two inclusive samples are merged before reweighting. The fractions f are determined from the inclusive sample. The fraction of $t\bar{t}$ events with 700 GeV $< M_{t\bar{t}} <$ 1000 GeV is 0.0967 and the fraction of events with $M_{t\bar{t}} >$ 1000 GeV is 0.0256. To cross-check the combined high statistics sample, certain kinematic distributions in the combined sample are compared against the kinematic distributions from the original inclusive sample. These comparisons are shown in Figure 6.15.

Sample	N_{events}	f
TT_TuneCUETP8M2T4_13TeV-powheg-pythia8	77229341	1.0
	78006311	1.0
TT_Mtt-700to1000_TuneCUETP8M2T4_13TeV-powheg-pythia8	38578334	0.0967
TT_Mtt-1000toInf_TuneCUETP8M2T4_13TeV-powheg-pythia8	24495211	0.0256

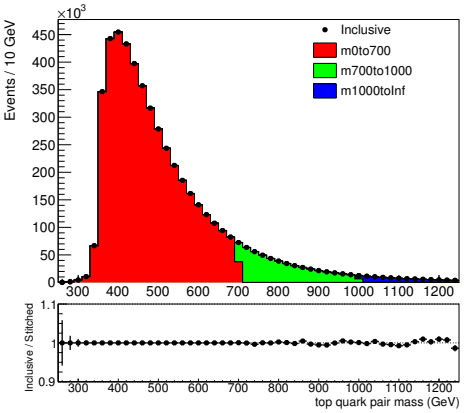
Table 6.16: MC samples used to improve statistics for the unfolding. The kinematic fit uses only the first part of the inclusive $t\bar{t}$ sample. The fraction f of events in each $M_{t\bar{t}}$ range is determined from the inclusive sample. The event counts and f for the inclusive samples describe the samples prior to the $M_{t\bar{t}} <$ 700 GeV requirement, applied later when combining the samples.

6.8.3 Response Matrix

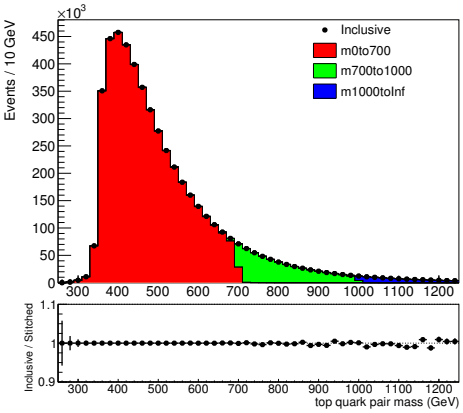
The $t\bar{t}$ simulation is used to produce a response matrix mapping the p_T of the t-tagged jet to the parton level top quark p_T, essentially parameterizing the de-



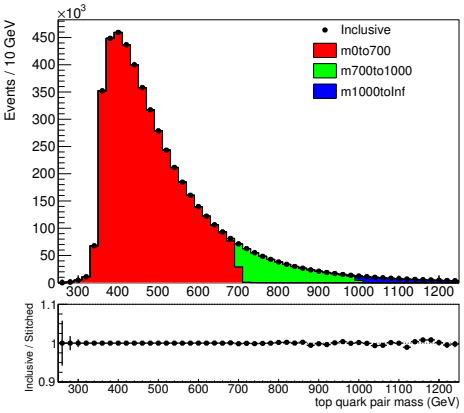
(a) μ+jets channel



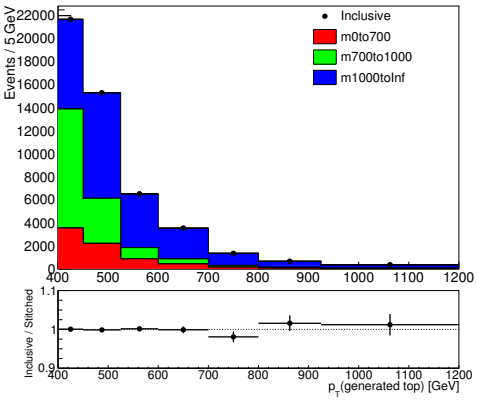
(b) e+jets channel



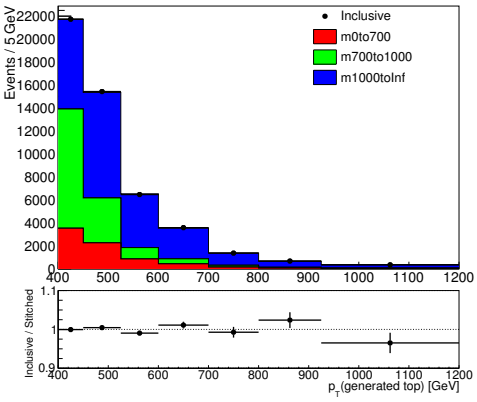
(c) μ+jets channel



(d) e+jets channel



(e) μ+jets channel

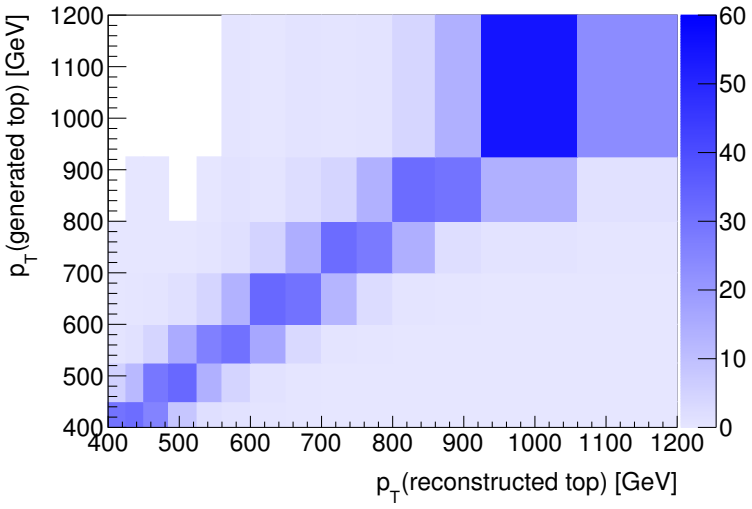


(f) e+jets channel

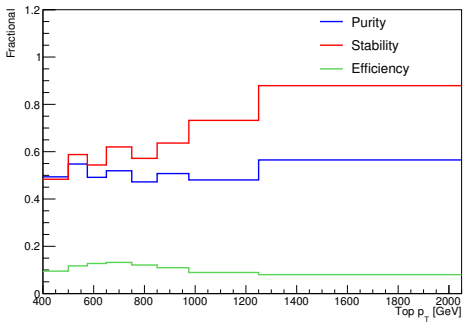
Figure 6.15: Comparison of kinematics between the inclusive and combined high statistics $t\bar{t}$ samples. The stitched label refers to the combined sample. Plots (a), (b) show $M_{t\bar{t}}$ for the hard process top quarks prior to FSR; (c), (d) show the parton level $M_{t\bar{t}}$ post-FSR; and (e), (f) show the parton level top quark p_T. The combined sample kinematics match the original inclusive sample, with improved statistics.

tector acceptance and efficiency. The unfolding procedure implicitly inverts this matrix, in order to extrapolate from the observed t-tagged jet p_T spectrum to the cross section as a function of top quark p_T. In order for this inversion to be well-defined, the response matrix must be reasonably smooth and diagonal. This is ensured through appropriate choices of the $\left(p_{\mathrm{T}}^{\mathrm{t\ jet}}, p_{\mathrm{T}}^{\mathrm{t}}\right)$ binning of the response matrix.

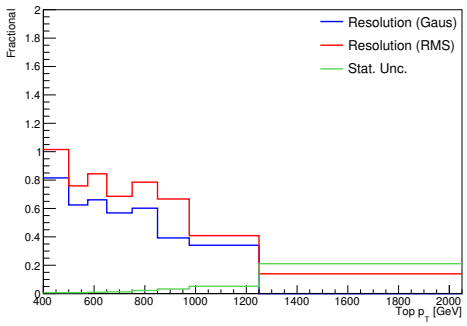
The scheme used to determine the response matrix binning assumes that the binning is the same for the t jet p_T and top quark p_T. First, the purity and stability in each bin is required to be generally greater than 50%. The purity is defined as the fraction of events in a given t jet p_T bin for which the top quark p_T is in the same bin, while the stability is the fraction of events in a given top quark p_T bin for which the t jet p_T is in the same bin. The bin width is also required to be smaller than the resolution, where the resolution is defined as either the RMS or the fitted Gaussian width of the t jet p_T for events in a given top quark p_T bin. In addition to the purity, stability, and resolution, the efficiency and relative statistical uncertainty are also considered when optimizing the binning. The efficiency is defined as the ratio of the number of events in a given t jet p_T bin to the number of events in the same top quark p_T bin, while the relative statistical uncertainty gives the ratio of the statistical uncertainty on the number of events in a given top quark p_T bin to the total number of events in that bin. Figures 6.16 and 6.17 show the response matrices using the optimized binning, along with the purity, stability, efficiency, resolution, and statistical uncertainty, for the μ+jets and e+jets channels respectively. While the binning optimization assumes the same binning in $p_T^{\mathrm{t\ jet}}$ and p_T^{t}, in practice each p_T^{t} bin is split into two $p_T^{\mathrm{t\ jet}}$ bins to satisfy a requirement of the unfolding algorithm.



(a) Response matrix

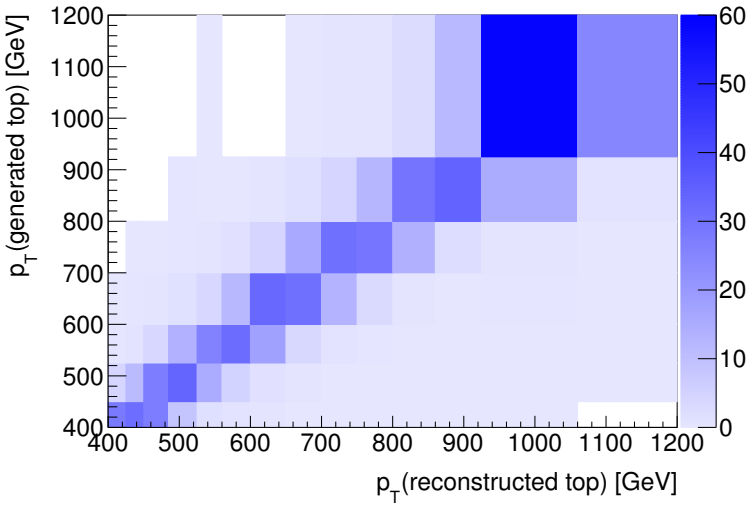


(b) Purity, stability, and efficiency

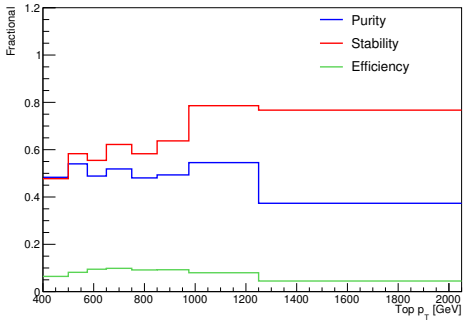


(c) Relative resolution and statistical uncertainty

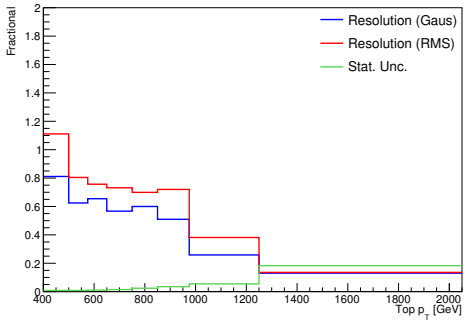
Figure 6.16: (a) Response matrix mapping the t jet p_T to the top quark p_T in the μ+jets channel, normalized to 100% in each bin of top quark p_T. (b) Stability, purity, and efficiency of each bin of the response matrix. (c) Relative resolution and statistical uncertainty in each bin of the response matrix. The relative resolution is defined as the resolution divided by the bin width, while the relative statistical uncertainty is the statistical uncertainty divided by the bin contents. The binning has been chosen such that the purity and stability are generally greater than 50% in all bins, and the RMS resolution is less than the bin width.



(a) Response matrix



(b) Purity, stability, and efficiency



(c) Relative resolution and statistical uncertainty

Figure 6.17: (a) Response matrix mapping the t jet p_T to the top quark p_T in the e+jets channel, normalized to 100% in each bin of top quark p_T. (b) Stability, purity, and efficiency of each bin of the response matrix. (c) Relative resolution and statistical uncertainty in each bin of the response matrix. The relative resolution is defined as the resolution divided by the bin width, while the relative statistical uncertainty is the statistical uncertainty divided by the bin contents. The binning has been chosen such that the purity and stability are generally greater than 50% in all bins, and the RMS resolution is less than the bin width.

6.8.4 TUnfold

The unfolding is performed using the TUnfold algorithm [92]. The algorithm is based on a least squares minimization with Tikhonov regularization. Specifically, the algorithm finds the stationary point of

$$
\begin{aligned}
\mathcal{L}(x,\lambda) &= \mathcal{L}_1 + \mathcal{L}_2 + \mathcal{L}_3 \\
\mathcal{L}_1 &= (y - Ax)^T V_{yy}^{-1} (y - Ax) \\
\mathcal{L}_2 &= \tau^2 (x - f_b x_0)^T (L^T L)(x - f_b x_0) \\
\mathcal{L}_3 &= \lambda (Y - e^T x) \\
Y &= \sum_i y_i \\
e_j &= \sum_i A_{ij}
\end{aligned}
\tag{6.9}
$$

Here, y is the observed result (with covariance matrix V_{yy}) and x is the unfolded measurement, with A the response matrix linking them. The term $\mathcal{L}_1$ describes the least-squares minimization, while $\mathcal{L}_2$ parameterizes the regularization. The regularization dampens fluctuations in x resulting from statistical fluctuations in y, but also biases the unfolding. The regularization is described by the regularization strength τ, the regularization mode L, and the bias x_0 and its strength f_b. $\mathcal{L}_3$ is an optional area constraint.

A number of studies were performed to determine the optimal unfolding procedure. Details on these studies may be found in [87]. From these studies a non-regularized unfolding was chosen to be optimal, setting $\tau = 0$. No area constraint was imposed. Therefore, the unfolding was performed as a simple least squares minimization, finding the result x which minimizes $\mathcal{L}_1$.

6.8.5 MC Studies

Before unfolding the observed t jet p_T distribution in data, closure tests of the unfolding procedure were performed to ensure that the unfolding was bias free. To perform the closure test, the $t\bar{t}$ MC sample was divided in half. One half of the sample was used to construct the response matrix, while the other half was used to construct a simulated t jet p_T distribution and top quark p_T distribution. The response matrix was then used to unfold the t jet p_T distribution, and the result compared to the top quark p_T distribution. Since the response matrix and the true relation between the simulated t jet p_T and top quark p_T distributions differ only by statistical fluctuations, the unfolded and generated top quark p_T distributions are expected to agree. Deviations between the unfolded and generated top quark p_T distributions therefore indicate biases in the unfolding or sensitivity to statistical fluctuations. To measure the unfolding bias and uncertainty, 1000 toy distributions are generated by randomly sampling the t jet p_T distribution. Each toy distribution is then unfolded and the result compared to the top quark p_T distribution. The mean difference between the distributions gives the unfolding bias, while the standard deviation describes the unfolding uncertainty.

To test the unfolding performance when the simulation used to produce the response matrix does not exactly match the conditions in data, the closure test is also performed with p_T reweighting. The events used to construct the t jet p_T and top quark p_T distributions are weighted by $(1.0 + 0.0004 \times p_T^{\mathrm{t}})^{\pm 1}$, coherently skewing the p_T distributions away from the response matrix kinematics. The reweighted measured distribution is then unfolded with the original response matrix, and compared to the reweighted truth distribution. As previously, the

unfolding is performed for 1000 toy distributions drawn from the reweighted t jet p_T distribution, with the results used to compute the unfolding bias and uncertainty. The closure test results are shown in Figures 6.18 and 6.19 for the μ+jets and e+jets channels, respectively. The unfolding shows a reasonable level of bias, and little dependence on the input t jet p_T shape.

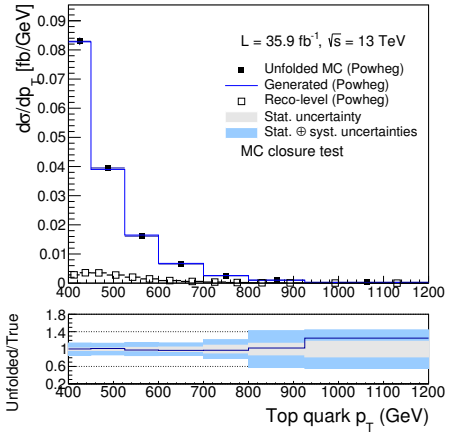
6.8.6 Differential Cross Section for Individual Channels

After performing the closure test to ensure that the unfolding is bias free, the unfolding procedure is applied to the measured t jet p_T distribution from data. The result is a measured parton level top quark p_T distribution. To construct the differential $t\bar{t}$ cross section as a function of the parton level top quark p_T, the unfolded distribution is corrected for the bin widths and scaled by the integrated luminosity and theoretical branching ratio for $t\bar{t} \to \ell\nu bq\bar{q}b$.

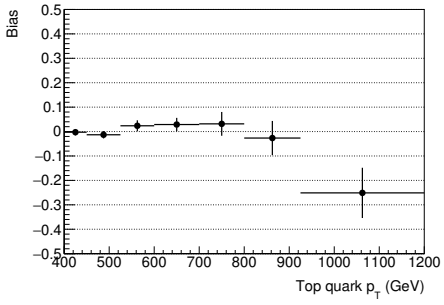
A number of systematic uncertainties, described in Section 6.6, affect the measured differential $t\bar{t}$ cross section. Other uncertainties on the unfolded result arise from the unfolding process itself. The impact of systematic uncertainties on the differential cross section is assessed by calculating a covariance matrix for the unfolded result x. For a non-regularized unfolding with no area constraint, the total covariance between bins of the unfolded result is given by

$$V_{xx} = \left(A^T V_{yy}^{-1} A\right)^{-1} + \sum_s V_{xx}^{\delta A_s} + V_{xx}^{\text{stat}}. \tag{6.10}$$

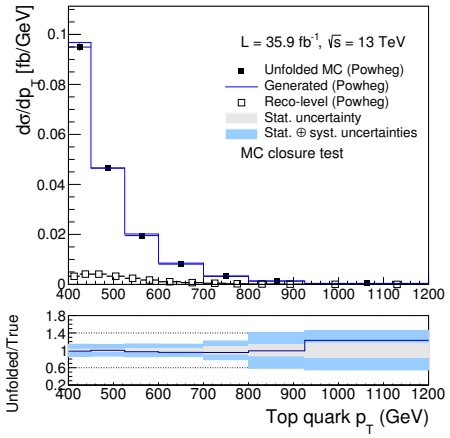
The first term describes contributions to the covariance on the unfolded result from the covariance on the observation V_{yy}. The observation covariance is



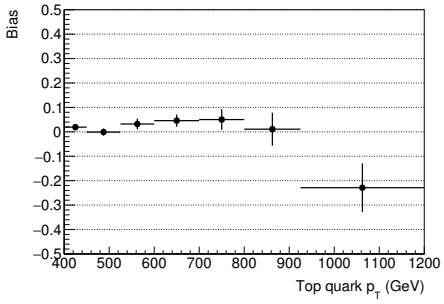
(a) closure test, nominal p_T



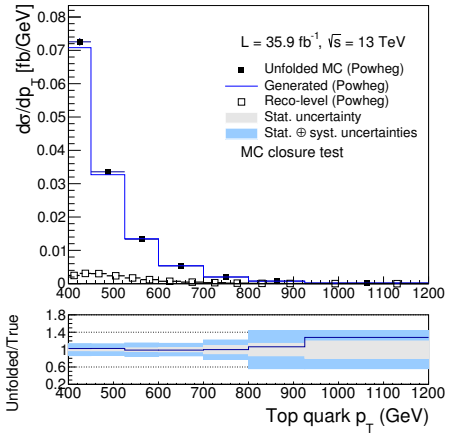
(b) bias, nominal p_T



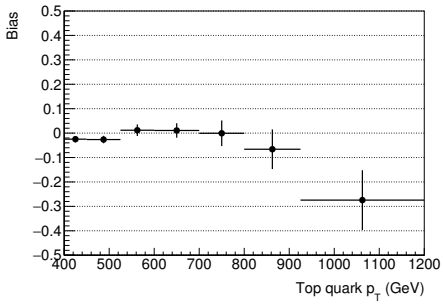
(c) closure test, p_T up



(d) bias, p_T up

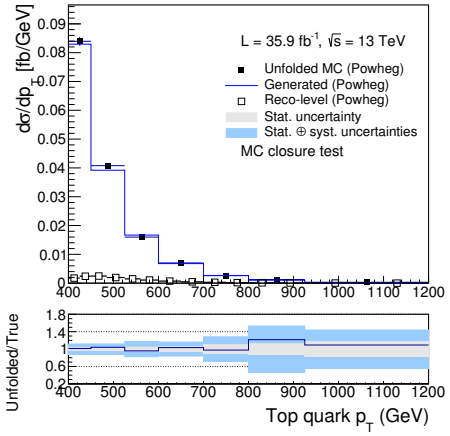


(e) closure test, p_T down

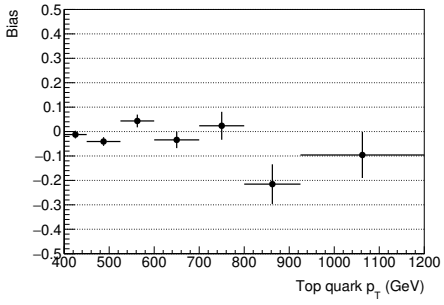


(f) bias, p_T down

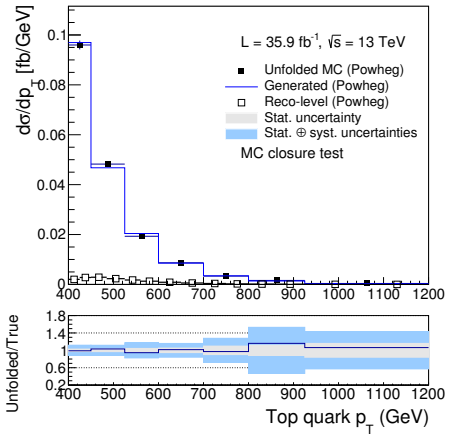
Figure 6.18: Closure test (left) and bias (right) for the unfolding in the μ+jets channel, for the nominal (top), p_T up (middle), and p_T down (bottom) distributions. The closure test unfolds half the MC $t\bar{t}$ sample with the other half and the bias describes the average difference between the unfolded and true top quark p_T distributions. The statistical and systematic uncertainties only include contributions from the MC statistics and unfolding systematics.



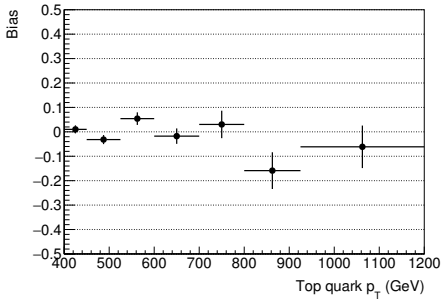
(a) closure test, nominal p_T



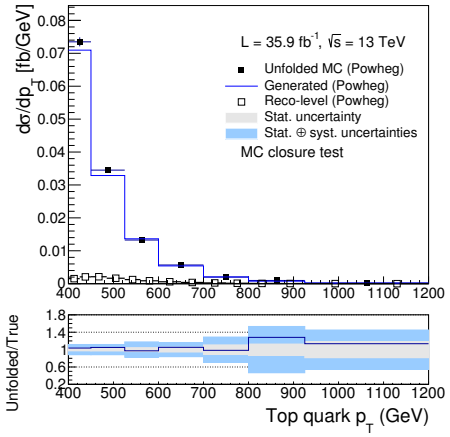
(b) bias, nominal p_T



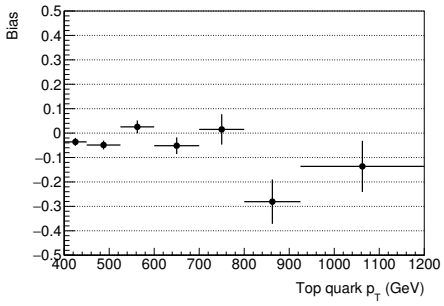
(c) closure test, p_T up



(d) bias, p_T up



(e) closure test, p_T down



(f) bias, p_T down

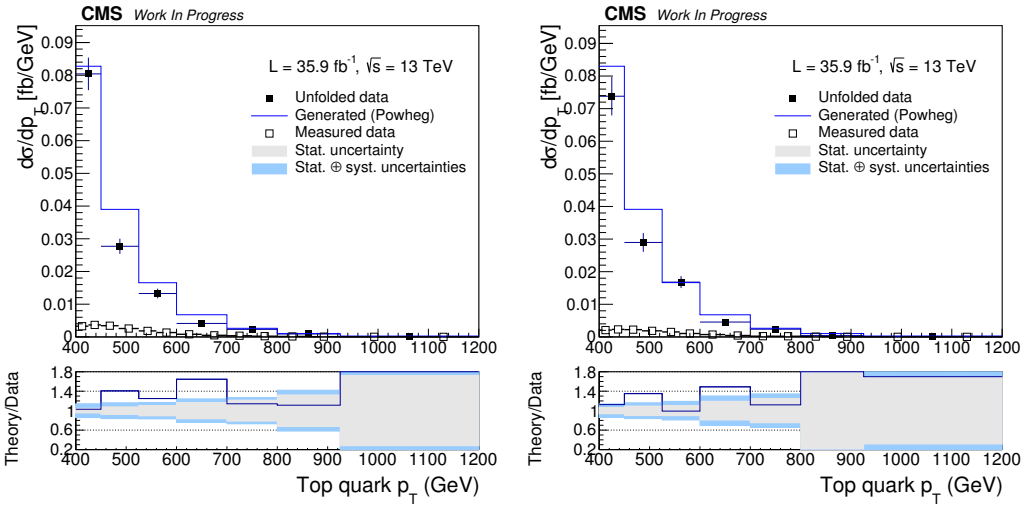
Figure 6.19: Closure test (left) and bias (right) for the unfolding in the e+jets channel, for the nominal (top), p_T up (middle), and p_T down (bottom) distributions. The closure test unfolds half the MC $t\bar{t}$ sample with the other half and the bias describes the average difference between the unfolded and true top quark p_T distributions. The statistical and systematic uncertainties only include contributions from the MC statistics and unfolding systematics.

propagated via the normalized response matrix A. V_{yy} includes statistical uncertainties on the data, statistical uncertainties on the simulated backgrounds subtracted from the data, and uncertainties on the background rates. The second term, $\sum_s V_{xx}^{\delta A_s}$, describes contributions to the covariance from systematic uncertainties affecting the response matrix. δA_s describes the impact of a 1σ systematic variation on the normalized response matrix A. The impact on the unfolded result x of an uncertainty δA_s on the response matrix can be determined through propagation of error, yielding an uncertainty δx on the unfolded result. The corresponding covariance matrix is then $V_{xx}^{\delta A} = \delta x\,(\delta x)^T$. Finally, the statistical uncertainty on the response matrix contributes a term V_{xx}^{stat} to the total covariance. More details on the calculation of the covariance matrix V_{xx} may be found in [92].

The measurements of the $t\bar{t}$ cross section as a function of top quark p_T performed in the μ+jets and e+jets channels are shown in Figure 6.20. The contributions to the uncertainty on the differential $t\bar{t}$ cross section resulting from different systematic uncertainties are shown in Figure 6.21.

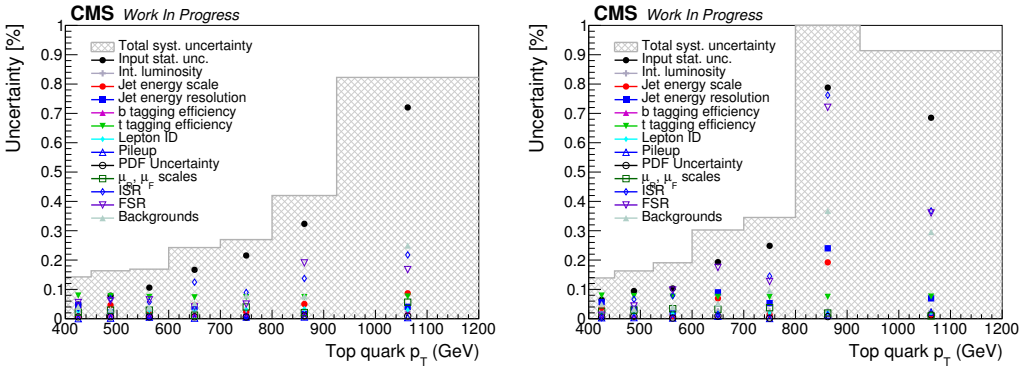
Figure 6.22 shows the total correlation matrices between bins of the unfolded result in the μ+jets and e+jets channels. The correlation matrix C_{xx} is related to the covariance matrix V_{xx} via:

$$C_{xx}^{ij} = \frac{V_{xx}^{ij}}{\sqrt{V_{xx}^{ii} * V_{xx}^{jj}}} \tag{6.11}$$



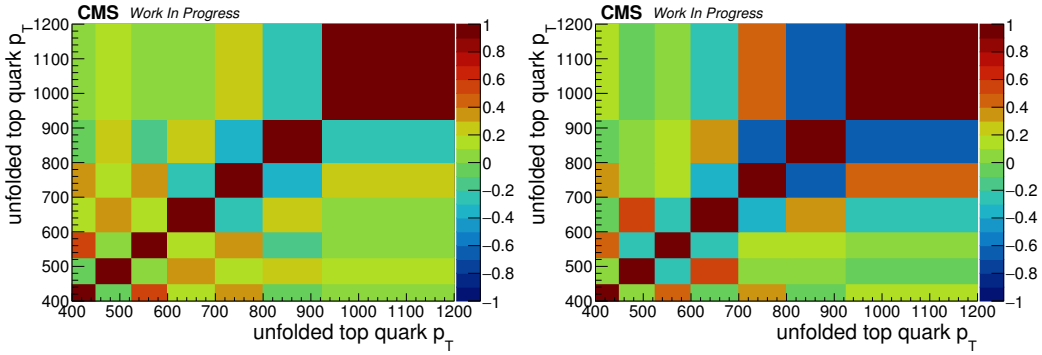
(a) μ+jets channel (b) e+jets channel

Figure 6.20: Parton level $t\bar{t}$ cross section as a function of top quark p_T, including all systematic uncertainties, as measured in the (a) μ+jets and (b) e+jets final states. The simulation predicts a harder top quark p_T spectrum than observed in data, though the two agree within uncertainties. Statistical uncertainty is the dominant uncertainty.



(a) μ+jets channel (b) e+jets channel

Figure 6.21: Contributions from various sources to the total uncertainty on the parton level $t\bar{t}$ cross section as a function of top quark p_T, for the cross section measurement in the (a) μ+jets and (b) e+jets final states. Statistical uncertainty is the dominant uncertainty, comprising statistical uncertainties on the data and response matrix. The main systematic uncertaities come from the choice of ISR and FSR scale, as well as the background rates.



(a) μ+jets channel (b) e+jets channel

Figure 6.22: Total correlation matrix between bins of the unfolded result in the (a) μ+jets and (b) e+jets channels. The correlation matrix is normalized to have a maximum correlation of 1.

6.8.7 Differential Cross Section for Combined Channels

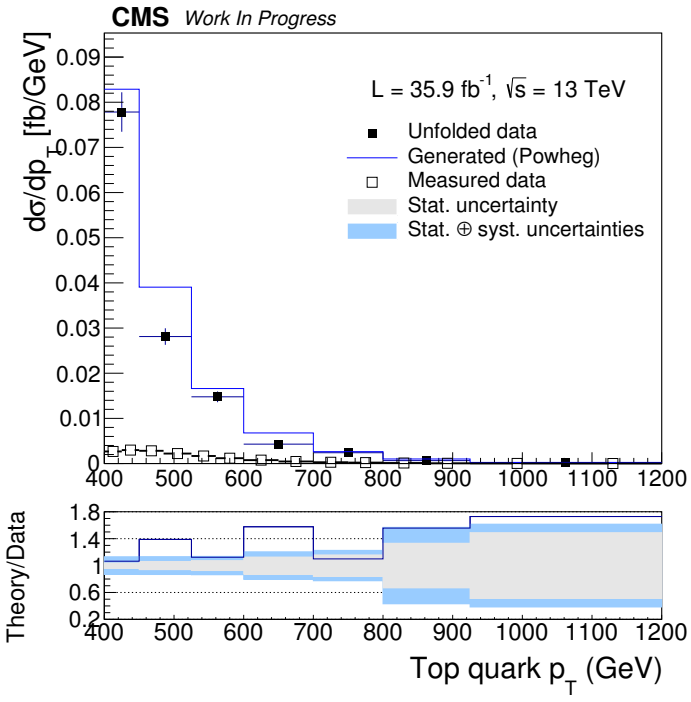
The final measurement of the differential $t\bar{t}$ cross section as a function of top quark p_T is performed combining the observations in the μ+jets and e+jets channels. A combined t jet p_T distribution is constructed using events from both channels, and a combined response matrix used to perform the unfolding. Systematic uncertainties on the combined result are handled in the same manner as previously. The combined data and background statistics are used to calculate V_{yy}, the response matrix uncertainties δA_s are calculated for the combined response matrix assuming the systematic to be fully correlated in both channels, and V_{xx}^{stat} is calculated for the combined response matrix. The improvement in statistics due to the combination reduces the uncertainties on the unfolded result.

The final parton level differential $t\bar{t}$ cross section as a function of top quark p_T, including all experimental and theoretical uncertainties, is shown in Figure 6.23. Figure 6.24 shows the covariance matrix for the final cross section. A

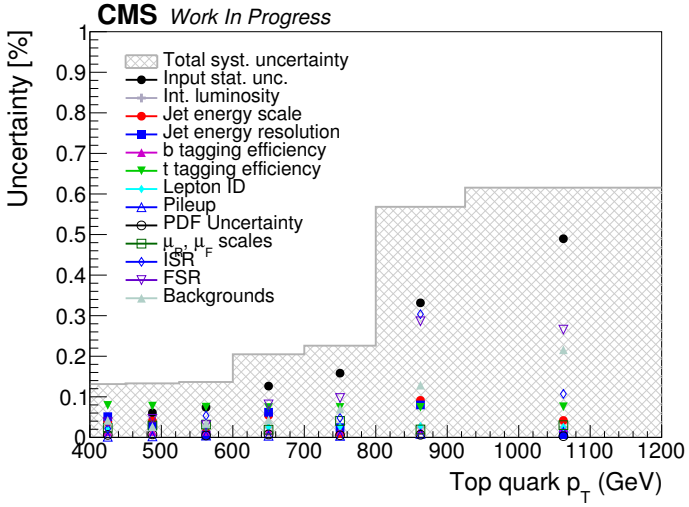
numerical description of the parton level differential $t\bar{t}$ cross section is given in Table 6.17, along with a numerical description of the different uncertainties.

$d\sigma/dp_T$ [fb/GeV] at parton level (combined result)							
p_T [GeV]	Data	Uncertainties [%]					POWHEG+PYTHIA8
		stat.	exp.	th.	lumi	total	
[400, 450]	77.83	3.86	11.28	4.93	2.50	13.14	82.88
[450, 525]	28.12	6.07	9.87	6.05	2.50	13.31	39.05
[525, 600]	14.78	7.40	8.80	6.92	2.50	13.65	16.61
[600, 700]	4.30	12.64	12.23	10.19	2.50	20.48	6.78
[700, 800]	2.40	15.83	10.96	11.57	2.50	22.60	2.63
[800, 925]	0.65	33.16	19.47	41.78	2.50	56.83	1.01
[925, 1200]	0.14	48.97	23.57	28.80	2.50	61.56	0.25

Table 6.17: Parton level $t\bar{t}$ cross section as a function of top quark p_T, presented in tabular form. The measured cross section is compared to the prediction from POWHEG+PYTHIA8 $t\bar{t}$ MC. The uncertainty on the measurement is divided into contributions from statistical, experimental, theoretical, and luminosity uncertainties.



(a)



(b)

Figure 6.23: Parton level $t\bar{t}$ cross section as a function of top quark p_T, along with breakdown of systematic uncertainties, from the measurement in the combined lepton+jets channel. The simulation predicts a harder top quark p_T spectrum than observed in data, though the two mostly agree within uncertainties. Statistical uncertainty is the dominant uncertainty, comprising statistical uncertainties on the data and response matrix. The main systematic uncertaities come from the choice of ISR and FSR scale, as well as the background rates.

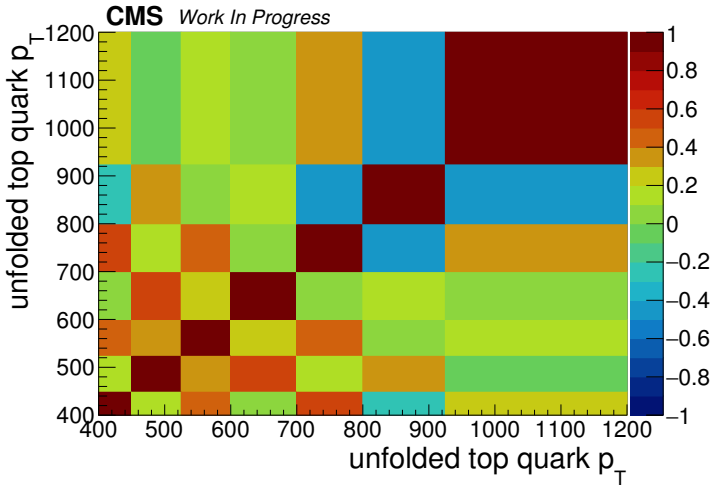


Figure 6.24: Total correlation matrix between bins of the parton level $t\bar{t}$ cross section as a function of the top quark p_T. The correlation matrix is normalized to have a maximum correlation of 100%.

6.9 Conclusion

This thesis presents a measurement of the $t\bar{t}$ cross section as a function of top quark p_T for top quarks with $p_T > 400$ GeV, where the $t\bar{t}$ pairs are produced in pp collisions with a center-of-mass energy of 13 TeV. The measurement is performed using data collected at the CMS detector during the 2016 LHC data-taking period, corresponding to an integrated luminosity of 35.9 fb^{-1}. The measurement is performed in the semileptonic final state, where one top quark decays hadronically via $t \to Wb \to q\bar{q}b$ and the other decays leptonically via $t \to Wb \to \ell\nu b$. At high energies, the jets formed by the hadronic top decay products merge, yielding a single wide t jet. Events in the boosted semileptonic final state therefore contain a t jet, b jet, muon or electron, and p_T^{miss}. These objects are reconstructed using data collected by the CMS detector.

To accurately record the information needed to reconstruct a particle collision, the CMS detector uses a series of subdetectors to measure specific particle types and properties. The CMS detector consists of a silicon tracker to record charged particle tracks, an electromagnetic calorimeter to measure the energies of charged particles, a hadronic calorimeter to measure the energy of hadrons, a 3.8 T superconducting solenoid to permit momentum measurements from charged particle tracks, and muon chambers to track muons. The electrons, muons, photons, neutral hadrons, and charged hadrons identified using these subdetectors are used to reconstruct higher-level objects such as jets or p_T^{miss}. The tracker is particularly important in reconstructing the displaced vertex used to identify b jets and the jet substructure used to identify t jets.

Events in the boosted semileptonic final state are used to construct an ob-

served t jet p_T distribution, which is unfolded to produce a measurement of the $t\bar{t}$ cross section as a function of top quark p_T. The observed t jet p_T distribution is constructed by subtracting expected background contributions from data in the boosted semileptonic final state. To account for differences between object identification and reconstruction efficiencies in simulation and data, SFs are applied to the simulated backgrounds prior to the subtraction. The t tag SF, as well as the background normalizations and lepton SFs, are determined through a kinematic fit to data. The fit is performed simultaneously in three kinematic regions differentiated by whether the b jet and t jet candidates in the event pass their respective tagging requirements, allowing the t tag SF to be isolated from the $t\bar{t}$ inclusive cross section. The remaining SFs are taken from independent measurements.

After constructing the background-subtracted data distribution, an unfolding procedure is used to extrapolate from the observed t jet p_T spectrum to the top quark p_T spectrum. The unfolding corrects for signal acceptance and bin migrations when reconstructing true semileptonic $t\bar{t}$ events. The semileptonic $t\bar{t}$ branching ratio is then used to extrapolate the measurement to the full parton-level phase space, and the integrated luminosity and bin widths used to calculate the differential cross section $d\sigma/dp_T$. The measured differential cross section is compared to the theory prediction from POWHEG+PYTHIA8 MC, which is observed to overestimate the data.

Measurements of the $t\bar{t}$ differential cross section provide important feedback on the accuracy of the simulation used to model $t\bar{t}$ events. Parameters such as α_S which control the simulation are tuned to reflect observations in data. As the LHC reach extends to higher energies, it becomes more important to

ensure simulations remain accurate in the high energy tail. Measurements in the boosted regime provide the information needed to tune simulation at high energies. Accurate modeling of the high energy tail is particularly important for top quark measurements, as many new physics models predict signatures with energetic top quarks.

APPENDIX A

LEPTON ISOLATION

This appendix describes the mini isolation algorithm used to identify signal leptons in a highly boosted regime. Mini isolation is similar to the standardized relative isolation prescription, in that it requires the sum p_T of the tracks in a cone around the lepton to be less than a given fraction of the lepton p_T. In contrast to relative isolation, the cone size is variable with respect to lepton p_T. Specifically, mini isolation is defined as

$$\text{miniIso} = \frac{1}{p_T^\ell} * [\text{CH} + \max(0, \text{NH} + \text{PH} - 0.5 * \text{PU} * \rho * \text{A}_{\text{eff}} * (\frac{\text{R}}{0.3})^2)] \tag{A.1}$$

where CH, NH, PH, and PU are the $\sum p_T$ of the charged hadrons, neutral hadrons, photons, and pileup particles, respectively, in a cone of inner width r and outer width R around the lepton. A_{eff} is the effective area, and ρ is the residual energy from pileup. R varies with the lepton p_T as

$$R = \begin{cases} 0.2 & p_T^\ell < 50 \text{ GeV} \\ \frac{10}{p_T^\ell} & 50 \text{ GeV} < p_T < 200 \text{ GeV} \\ 0.05 & p_T^\ell > 200 \text{ GeV} \end{cases} \tag{A.2}$$

The inner cone width r is dependent on the lepton and the isolation constituent, and is defined in Table A.1.

Constituent	Muon	Electron (barrel)	Electron (endcap)
NH	0.0001	0.0	0.015
CH	0.01	0.0	0.0
PH	0.01	0.0	0.08
PU	0.01	0.0	0.015

Table A.1: Inner radii of the cone around the lepton in mini isolation. Constituents are not added to the mini isolation sum if they are separated from the lepton by less than this amount.

BIBLIOGRAPHY

[1] Y. Nir, "The Standard Model." https://indico.cern.ch/event/353089/contributions/1762251/attachments/699449/960335/sm_nir.pdf. Presented at the 2015 CERN-Fermilab HCP Summer School, CERN, 24-26 June 2015.

[2] P. Langacker, *The Standard Model and Beyond*. Taylor & Francis, Boca Raton, FL, USA, 2010.

[3] S. Martin, *A Supersymmetry Primer*, [hep-ph/9709356].

[4] S.-F. Ge, H.-J. He and R.-Q. Xiao, *Probing New Physics Scales from Higgs and Electroweak Observables at e^+e Higgs Factory*, *JHEP* **10** (2016) 007, [hep-ph/1603.03385].

[5] G. Farrar and P. Fayet, *Phenomenology of the production, decay, and detection of new hadronic states associated with supersymmetry*, *Phys. Lett. B* **76** (1978) 575.

[6] B. Lillie, L. Randall and L.-T. Wang, *The Bulk RS KK-gluon at the LHC*, *JHEP* **09** (2007) 074, [hep-ph/0701166v1].

[7] K. Agashe, A. Belyaev, T. Krupovnickas, G. Perez and J. Virzi, *LHC Signals from Warped Extra Dimensions*, *Phys. Rev. D* **77** (2008) 015003, [hep-ph/0612015v1].

[8] G. Shiber, "Randall-Sundrum Cosmology and Dp-Brane Dynamics." https://www.georgeshiber.com/randall-sundrum-cosmology-and-dp-brane-dynamics/.

[9] Particle Data Group, J. Beringer et al., *Review of particle physics*, *Phys. Rev. D* **86** (2012) 010001.

[10] P. Skands, "QCD and Monte Carlos." https://indico.cern.ch/event/353089/contributions/1762257/attachments/699461/960348/HCPSS-1.pdf. Presented at the 2015 CERN-Fermilab HCP Summer School, CERN, 24-26 June 2015.

[11] F. Kohn, *Measurement of the charge asymmetry in top quark pair production in pp collision data at $\sqrt{s} = 7$ TeV using the ATLAS detector*, Ph.D. thesis, U. Gottingen (main), 2012, [hep-ex/1204.0952].

[12] M. Czakon, D. Heymes, A. Mitov, D. Pagani, I. Tsinikos and M. Zaro, *Top-pair production at the LHC through NNLO QCD and NLO EW*, *JHEP* **10** (2017) 186, [hep-ph/1705.04105].

[13] P. Nason, *A new method for combining NLO QCD with shower Monte Carlo algorithms*, *JHEP* **11** (2004) 040, [hep-ph/0409146].

[14] S. Frixione, P. Nason and C. Oleari, *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, *JHEP* **11** (2007) 070, [hep-ph/0709.2092].

[15] S. Alioli, P. Nason, C. Oleari and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, *JHEP* **06** (2010) 043, [hep-ph/1002.2581].

[16] J. M. Campbell, R. K. Ellis, P. Nason and E. Re, *Top-pair production and decay at NLO matched with parton showers*, *JHEP* **04** (2015) 114, [hep-ph/1412.1828].

[17] T. Sjöstrand, S. Mrenna and P. Skands, *PYTHIA 6.4 Physics and Manual*, *JHEP* **05** (2006) 026, [hep-ph/0603175].

[18] T. Sjöstrand, S. Mrenna and P. Skands, *A Brief Introduction to PYTHIA 8.1*, *Comput. Phys. Commun.* **178** (2008) 852, [hep-ph/0710.3820].

[19] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten et al., *An Introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) 159, [hep-ph/1410.3012].

[20] P. Skands, S. Carrazza and J. Rojo, *Tuning PYTHIA 8.1: the Monash 2013 Tune*, *Eur. Phys. J. C* **74** (2014) 3024, [hep-ph/1404.5630].

[21] CMS Collaboration, *Investigations of the impact of the parton shower tuning in Pythia 8 in the modelling of $t\bar{t}$ at $\sqrt{s} = 8$ and 13 TeV*, CMS Physics Analysis Summary CMS-PAS-TOP-16-021, 2016.

[22] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, *JHEP* **07** (2014) 079, [hep-ph/1405.0301].

[23] CMS Collaboration, V. Khachatryan et al., *Event generator tunes obtained*

from underlying event and multiparton scattering measurements, *Eur. Phys. J. C* **76** (2016) 155, [hep-ex/1512.00815].

[24] J. Alwall, S. Hoeche, F. Krauss, N. Lavesson, L. Lonnblad, F. Maltoni et al., *Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions*, *Eur. Phys. J. C* **53** (2008) 473, [hep-ph/0706.2569].

[25] R. Frederix and S. Frixione, *Merging meets matching in MC@NLO*, *JHEP* **12** (2012) 061, [hep-ph/1209.6215].

[26] CMS Collaboration, A. M. Sirunyan et al., *Search for t t-bar resonances in highly boosted lepton+jets and fully hadronic final states in proton-proton collisions at sqrt(s) = 13 TeV*, *JHEP* **07** (2017) 001, [hep-ex/1704.03366].

[27] CMS Collaboration, *Search for direct top squark pair production in the dilepton final state at* $\sqrt{s}$ = 13 *TeV*, CMS Physics Analysis Summary CMS-PAS-SUS-16-027, 2016.

[28] *ECFA-CERN Workshop on Large Hadron Collider in the LEP tunnel*, Lausanne, Switzerland, March 1984, DOI.

[29] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, *Phys. Lett. B* **716** (2012) 1, [hep-ex/1207.7214].

[30] CERN Collaboration, E. Mobs, *The CERN accelerator complex. Complexe des accélérateurs du CERN*, OPEN-PHO-ACCEL-2016-009, 2016.

[31] L. Evans and P. Bryant, *LHC machine*, *JINST* **3** (2008) S08001.

[32] M. Lamont, "The LHC machine – present and future." https://indico.cern.ch/event/226365/contributions/1533492/attachments/372511/518303/LHC-present-future-lamont-part1.pdf. Presented at the 2013 CERN-Fermilab HCP Summer School, CERN, 31 August 2013.

[33] "CMS Luminosity – Public Results." https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults.

[34] "Performance 2016." https://lhc-commissioning.web.cern.ch/lhc-commissioning/performance/2016-performance.htm.

[35] CMS Collaboration, S. Chatrchyan et al., *The CMS Experiment at the CERN LHC*, *JINST* **3** (2008) S08004.

[36] T. Sakuma and T. McCauley, *Detector and Event Visualization with SketchUp at the CMS Experiment*, *J. Phys. Conf. Ser.* **513** (2014) 002032, [physics.ins-det/1512.00815].

[37] CMS Collaboration, D. Barney, *CMS Detector Slice*, CMS-PHO-GEN-2016-001, 2016.

[38] CMS Collaboration, *CMS: The TriDAS project. Technical design report, Vol. 2: Data acquisition and high-level trigger*, CERN-LHCC-2002-026, 2002.

[39] "CMSSW Application Framework." https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookCMSSWFramework.

[40] CMS Collaboration, *CMS Tracker Technical Design Report*, CERN/LHC 98-6, 1998.

[41] W. Erdmann, *The CMS pixel detector*, *Int. J. Mod. Phys.* **A25** (2010) 1315.

[42] CMS Collaboration, *Description and performance of track and primary-vertex reconstruction with the CMS tracker*, *JINST* **9** (2014) P10009, [physics.ins-det/1405.6569].

[43] CMS Collaboration, A. M. Sirunyan et al., *Particle-flow reconstruction and global event description with the CMS detector*, *JINST* **12** (2017) P10003, [physics.ins-det/1706.04965].

[44] CMS Collaboration, *The performance of the CMS muon detector in proton-proton collisions at* $\sqrt{s} = 7$ *TeV at the LHC*, *JINST* **8** (2013) P11002, [physics.ins-det/1306.6905].

[45] CMS Collaboration, *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software*, CERN-LHCC-2006-001, CMS-TDR-8-1, 2006.

[46] CMS Collaboration, *Pileup Removal Algorithms*, CMS Physics Analysis Summary CMS-PAS-JME-14-001, 2014.

[47] W. Adam, R. Frühwirth, A. Strandlie and T. Todorov, *Reconstruction of*

electrons with the Gaussian-sum filter in the CMS tracker at the LHC, *J. Phys. G Nucl. Part. Phys.* **31** (2005) N9.

[48] CMS Collaboration, V. Khachatryan et al., *Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at* $\sqrt{s}$ = 8 *TeV*, *JINST* **10** (2015) P06005, [physics.ins-det/1502.02701].

[49] I. Kravchenko, "Update on 80X electron cut-based ID." https://indico.cern.ch/event/482677/contributions/2259342/attachments/1316731/1972911/talk_electron_ID_spring16_update.pdf. Presented at Joint ECAL DPG/EGM POG meeting, CERN, 26 July 2016.

[50] K. Rehermann and B. Tweedie, *Efficient identification of boosted semileptonic top quarks at the LHC*, *JHEP* **03** (2011) 059, [hep-ph/1007.2221].

[51] "SUSY Leptons selection and Data/MC Scale Factors." https://twiki.cern.ch/twiki/bin/viewauth/CMS/SUSLeptonSF.

[52] CMS Collaboration, S. Chatrchyan et al., *Performance of CMS muon reconstruction in pp collision events at* $\sqrt{s}$ = 7 *TeV*, *JINST* **7** (2012) P10002, [physics.ins-det/1206.4071].

[53] C. Botta and G. Petrucciani, "Study of present and new muon ID selection criteria." https://indico.cern.ch/event/357213/contributions/1769745/attachments/710701/975626/muonid-pog081214.pdf. Presented at Muon POG meeting, CERN, 8 December 2014.

[54] "Work in progress (evolving results) and dedicated studies from PAGs." https://twiki.cern.ch/twiki/bin/view/CMS/MuonWorkInProgressAndPagResults.

[55] M. Cacciari, G. P. Salam and G. Soyez, *The anti-k_t jet clustering algorithm*, *JHEP* **04** (2008) 063, [hep-ph/0802.1189].

[56] M. Cacciari, G. P. Salam and G. Soyez, *FastJet User Manual*, *Eur. Phys. J. C* **72** (2012) 1896, [hep-ph/1111.6097].

[57] CMS Collaboration, *Jet algorithms performance in 13 TeV data*, CMS Physics Analysis Summary CMS-PAS-JME-16-003, 2016.

[58] CMS Collaboration, V. Khachatryan et al., *Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV*, *JINST* **12** (2017) P02014, [hep-ex/1607.03663].

[59] "Recommended Jet Energy Corrections and Uncertainties For Data and MC."
https://twiki.cern.ch/twiki/bin/view/CMS/JECDataMC.

[60] CMS Collaboration, *Identification of b quark jets at the CMS experiment in the LHC Run 2*, CMS Physics Analysis Summary CMS-PAS-BTV-15-001, 2016.

[61] "Usage of b/c Tag Objects for 13 TeV Data in 2016 and 80X MC."
https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagRecommendation80XReReco.

[62] M. Dasgupta, A. Fregoso, S. Marzani and G. Salam, *Towards an understanding of jet substructure*, *JHEP* **09** (2013) 029, [hep-ph/1307.0007].

[63] A. Larkoski, S. Marzani, G. Soyez and J. Thaler, *Soft Drop*, *JHEP* **05** (2014) 146, [hep-ph/1402.2657].

[64] Y. Dokshitzer, G. Leder, S. Moretti and B. Webber, *Better Jet Clustering Algorithms*, *JHEP* **08** (1997) 001, [hep-ph/9707323].

[65] M. Wobisch and T. Wengler, *Hadronization corrections to jet cross sections in deep- inelastic scattering*, [hep-ph/9907280].

[66] J. Thaler and K. Van Tilburg, *Identifying boosted objects with N-subjettiness*, *JHEP* **03** (2011) 015, [hep-ph/1011.2268].

[67] J. Thaler and K. Van Tilburg, *Maximizing boosted top identification by minimizing N-subjettiness*, *JHEP* **02** (2012) 093, [hep-ph/1108.2701].

[68] S. Catani, Y. Dokshitzer, M. Seymour and B. Webber, *Longitudinally-invariant $k_\perp$-clustering algorithms for hadron-hadron collisions*, *Nucl. Phys. B* **406** (1993) 187.

[69] S. Ellis and D. Soper, *Successive Combination Jet Algorithm for Hadron Collisions*, *Phys. Rev. D* **48** (1993) 3160, [hep-ph/9305266].

[70] Particle Data Group, C. Patrignani et al., *Review of particle physics*, *Chin. Phys. C* **40** (2016) 100001.

[71] CMS Collaboration, V. Khachatryan et al., *Measurement of differential cross sections for top quark pair production using the lepton+jets final state in proton-proton collisions at 13 TeV*, *Phys. Rev. D* **95** (2017) 092001, [hep-ex/1610.04191].

[72] CMS Collaboration, V. Khachatryan et al., *Measurement of the integrated and differential t-tbar production cross sections for high-pt top quarks in pp collisions at $\sqrt{s}$ = 8 TeV*, *Phys. Rev. D* **94** (2016) 072002, [hep-ex/1605.00116].

[73] ATLAS Collaboration, G. Aad et al., *Measurement of the differential cross-section of highly boosted top quarks as a function of their transverse momentum in $\sqrt{s}$ = 8 TeV proton-proton collisions using the ATLAS detector*, *Phys. Rev. D* **93** (2016) 032009, [hep-ex/1510.03818].

[74] CMS Collaboration, *Measurement of the $t\bar{t}$ production cross section at 13 TeV in the all-jets final state*, CMS Physics Analysis Summary CMS-PAS-TOP-16-013, 2016.

[75] ATLAS Collaboration, *Measurements of top-quark pair differential cross-sections in the lepton+jets channel in pp collisions at $\sqrt{s}$ = 13 TeV using the ATLAS detector*, CERN-EP-2017-058, 2017.

[76] NNPDF Collaboration, R. D. Ball et al., *Parton distributions for the LHC Run II*, *JHEP* **04** (2015) 040, [hep-ph/1410.8849].

[77] M. Czakon and A. Mitov, *Top++: a program for the calculation of the top-pair cross-section at hadron colliders*, *Comput. Phys. Commun.* **185** (2014) 2930, [hep-ph/1112.5675].

[78] S. Agostinelli et al., *Geant4-a simulation toolkit*, *Nucl. Instr. Meth. A* **506** (2003) 250.

[79] P. Kant, O. Kind, T. Kintscher, T. Lohse, T. Martini, S. Mölbitz et al., *HatHor for single top-quark production: Updated predictions and uncertainty estimates for single top-quark production in hadronic collisions*, *Comput. Phys. Commun.* **191** (2015) 74, [hep-ph/1406.4403].

[80] N. Kidonakis, *NNLL threshold resummation for top-pair and single-top production*, *Phys. Part. Nucl.* **45** (2014) 714, [hep-ph/1210.7813].

[81] Y. Li and F. Petriello, *Combining QCD and electroweak corrections to dilepton production in the framework of the FEWZ simulation code*, *Phys. Rev. D* **86** (2012) 094034, [hep-ph/1208.5967].

[82] "Summary table of samples produced for the 1 Billion campaign, with 25ns bunch-crossing." https://twiki.cern.ch/twiki/bin/view/CMS/SummaryTable1G25ns.

[83] T. Gehrmann, M. Grazzini, S. Kallweit, P. Maierhöfer, A. von Manteuffel, S. Pozzorini et al., *W^+W^- Production at Hadron Colliders in Next to Next to Leading Order QCD*, *Phys. Rev. Lett.* **113** (2014) 212001, [hep-ph/1408.5243].

[84] F. Cascioli, T. Gehrmann, M. Grazzini, S. Kallweit, P. Maierhfer, A. von Manteuffel et al., *ZZ production at hadron colliders in NNLO QCD*, *Phys. Lett. B* **735** (2014) 311, [hep-ph/1405.2219].

[85] J. M. Campbell, R. K. Ellis and C. Williams, *Vector boson pair production at the LHC*, *JHEP* **07** (2011) 018, [hep-ph/1105.0020].

[86] "Top-tagging WPs 80X." https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetTopTagging.

[87] S. Dittmer, L. Skinnari and J. Thom, *Measurement of the differential $t\bar{t}$ production cross section for high-p_T top quarks in e/μ+jets final states at 13 TeV*, CMS AN-16-174.

[88] CMS Collaboration, *CMS Luminosity Measurements for the 2016 Data Taking Period*, CMS Physics Analysis Summary CMS-PAS-LUM-17-001, 2017.

[89] "Top Systematic Uncertainties (Run 2)." https://twiki.cern.ch/twiki/bin/viewauth/CMS/TopSystematics.

[90] J. Bendavid, "New MC Features for Analysis in 74x." https://indico.cern.ch/event/459797/contributions/1961581/attachments/1181555/1800214/mcaod-Feb15-2016.pdf. Presented at the PPD General Meeting, CERN, 4 November 2015; updated 15 February 2016.

[91] ATLAS and CMS Collaborations, LHC Higgs Combination Group, *Procedure for the LHC Higgs boson search combination in Summer 2011*, ATL-PHYS-PUB 2011-11, CMS-NOTE-2011-005, 2011.

[92] S. Schmitt, *TUnfold, an algorithm for correcting migration effects in high energy physics*, *JINST* **7** (2012) T10003, [physics.data-an/1205.6201v4].